THE ROLE OF LANGUAGE IN LEARNING, COMMUNICATING, AND DOING CHEMISTRY

JACKY DENG

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctorate in Philosophy in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

July 2024
© Jacky Deng, Ottawa, Canada, 2024

Abstract

We live in a world facing a myriad of complex global challenges that will require all citizens to be able to make and justify decisions using scientific evidence. We also live in an increasingly global society, one in which linguistic diversity in chemistry classrooms and research spaces is prevalent and increasing. If we are to succeed in solving the challenges facing our world, we must learn how to equip students with the scientific skills required to navigate these challenges, while also prioritizing equity so that all students have opportunities to learn, regardless of their language background. The absence of research on these topics risks impeding progress in scientific literacy and inclusion, limiting our capacity to navigate and solve pressing global issues. Using a combination of qualitative and quantitative research methods, my doctoral work has been guided by two overarching questions in chemistry education research: (1) how do students engage in argumentation and reasoning in chemistry education? and (2) how do students from diverse language backgrounds learn, communicate, and do chemistry?

My research has uncovered that how students construct arguments about chemical phenomena—including the reasoning, granularity, and comparisons in arguments—can depend on different curricular factors and expectations (Chapters 2 and 3). Implications from this work include how constructive alignment and backwards design can be used by chemistry educators to support the development of students' reasoning and argumentation skills across a curriculum (Chapter 4). In addition to curricular factors, my research has studied the role of students' language backgrounds in chemistry learning and research. We have found that English proficiency and history is associated with how students justify chemical phenomena, suggesting a need to identify the barriers encountered by students from diverse language backgrounds in chemistry classrooms to ensure all students have equitable ways to demonstrate their abilities (Chapter 5). We found that Eng+ chemistry trainees face pervasive linguistic challenges and barriers in research, but also resilience and a growth mindset, with implications for how supervisors and institutions can better support them (Chapters 6 and 7). My doctoral thesis also includes work I have led or contributed to on developing and refining chemistry education research instruments (Chapter 8), demonstrating the impacts of chemistry graduate student initiatives on advancing equity and inclusion (Chapter 9), surveying the state graduate student finances across Canada, including identifying challenges and solutions (Chapter 10), and highlighting ways to engage in effective and equitable mentorship strategies in chemistry and science, including potential impacts (Chapter 11). Collectively, the work described in my thesis has contributed to many areas of chemistry education research, with broad implications for supporting students reasoning in chemistry, linguistic equity in chemistry, instrument development in chemistry education research, and chemistry graduate student education.

Acknowledgements

Everything in this thesis was conducted on the traditional unceded territory of the Algonquin-Anishinaabe. The Algonquin-Anishinaabe are one of Canada's largest First Nations communities. Prior to colonization, their territory included the northern shores of Lake Huron and Lake Superior to the Canadian prairies. Since being colonized, the Algonquin-Anishinaabe have fought to live and retain the rights to their land through treaties and negotiations with the federal government of Canada and provincial governments of Ontario and Québec. My acknowledgement is intended to convey my explicit support for their land rights, as well as the paradoxical and unfortunate fact that this thesis would not be possible without colonization of their land. As such, I believe that my acknowledgement is both important to include, but also profoundly privileged and inherently limited in contributing to any meaningful change. I encourage all readers of this thesis to learn more about the history of colonization and genocide in Canada, as well as actionable ways to support Indigenous communities across Canada (such as those described in the Truth and Reconciliation Commission of Canada).

The work described in this thesis was funded by the Government of Ontario through an Ontario Graduate Scholarship (OGS), the Natural Science and Engineering Research Council (NSERC) through a Canadian Graduate Scholarship-Master's (CGS-M) Scholarship, and the Social Science and Humanities Research Council (SSHRC) through a Vanier Graduate Scholarship.

To the members of my Thesis Advisory Committee (Prof. Vicente Talanquer, Prof. Giuliano Reis, and Prof. Jeffrey Keillor): thank you to each of you for your time and thoughtful, critical, and constructive feedback over the years. Prof. Talanquer: You have been an inspiration to me since I first developed an interest in chemistry education, and it has been an absolute privilege to have conversations with you about my research and the future of the field. I hope we can have more in the future. Prof. Reis: Thank you for bearing with a group of chemists for five years. Your perspectives as both an educator and an educational researcher have been invaluable for my research and learning. Thank you for always sharing literature and feedback to support me. Prof. Keillor: You were the first professor to make me feel at home at uOttawa

and one of the first people who made me feel like I belonged in graduate school. In addition to the feedback and questions you have shared over the years, I'll forever appreciate your kindness, humility, and sense of humour.

Thank you to the various mentors and colleagues I've had the privilege of learning from and working with over the years. Everyone on the various committees and groups I've been a part of—the Working for Inclusion, Diversity, and Equity (WIDE) Committee, the Chemistry Education Division Virtual Seminar Series Working Group, the Society of Teaching and Learning in Higher Education's (STLHE) EDI Task Force, the uOttawa Chemistry and Biomolecular Science's Graduate Students' Association (CBGSA), Canadians Working for Inclusivity in the Chemical Sciences (CWIC), Ottawa Science Policy Network (OSPN), and Support Our Science—thank you for your commitment to learning and making chemistry, science, and academia a kinder, more equitable place. Prof. Fred Menard for showing me the beautiful art of organic chemistry and for always believing in me. Prof. W. Stephen McNeil for taking a chance on me as an undergrad, inspiring me to pursue a graduate degree in chemistry education, and being my biggest cheerleader every step of the way.

To the Flynn Research Group (FRG): I could not think of a better team of people to work with and am thankful to each one of you for the friendships and memories we have created over the years. Thank you (in no particular order): Alisha, Amanda, Keith, Myriam, Emily, Nick, Ahmed, Zena, Samira, Denzel, Lauren, Zahra, Kathryn, Kevin, Elly, Malek, Matthew, Allison, Sarah, Hossein, Uche, and Silvi. To the future FRG member who happens to read this: I hope this document is helpful in some way and please feel free to reach out if you need anything! Remember to breathe and have fun!

Mom and Dad, 感谢你们漂洋过海，放下自己的梦想和愿望，支持我们的工作. Kitty and Myra, thank you for keeping me grounded and being my two thousand friends. I love each of you dearly. Even though you'll never read it all, I dedicate this thesis to you.

Thank you to my friends. To the many runners and basketball players who helped dissipate any stress I had with a long run or game of pick-up. Riley, for those hilarious videos and memories we made together as roommates—I wouldn't be doing this if it wasn't for our shared anxiety and joy as chemists and educators. Brent and Simon for the wonderful memories we shared in Montréal. Garett, for always discussing the trolley problem with me (we'll figure it out one day!). Vince and Lindsay, for being the best running friends I could ask for (see you at Race Weekend!). Isabelle, for your warmth. Jenna and Dave, for letting me (a straight guy!) hang out with you. Billy, Keith, and Haleigh, for our nights as the Brew Crew. Adam, for your joy, authenticity, and empathy—you inspire me to be a better person, friend, and academic. Ravi, for reminding me that, at the end of the day, Surrey will always be a part of me.

Maria: thank you for the beautiful life we have built and continue to build. None of this would be possible without your infinite support, wisdom, joy, and love. You inspire me to a better man, human, and scientist. I can't wait for the rest of our lives together.

Prof. Alison Flynn: thank you for everything. I will forever admire your leadership, drive, and grace. You gave me the confidence and freedom to grow and learn. Thank you for always pushing me to ask what's important—in research and beyond.

Table of Contents

Abstract ii
Acknowledgements iii
List of Figures x
List of Tables xi
List of Abbreviations xix
Chapter 1 : Introduction 1
1.1 Introduction to Chemistry Education Research (CER) 1
1.1.1 What is chemistry education research and why is it important? 1
1.1.2 How is chemistry education research conducted? 3
1.1.2.1 Research questions in chemistry education research 3
1.1.2.2 Theoretical frameworks in chemistry education research 3
1.1.2.3 Ethics in chemistry education research 4
1.1.2.4 Qualitative and quantitative methods in chemistry education research 4
1.1.2.5 Instruments in chemistry education research 5
1.1.3 Knowledge mobilization 12
1.1.4 Conclusion 13
1.2 Argumentation in chemistry education 13
1.2.1 Reasoning in chemistry education 15
1.2.2 Granularity in science education 18
1.3 Linguistic equity in chemistry education and research 20
1.3.1 Linguistic equity and scientific argumentation 23
1.3.2 Linguistic equity in chemistry research 25
1.4 Thesis overview 27
1.5 References for Chapter 1 28
Chapter 2 : Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms 51
2.1 Introduction to the research article presented in this chapter 51
2.2 References for Introduction 52
2.3 Copyright 52
Chapter 3 : Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items 68
3.1 Introduction to the research article presented in this chapter 68

3.2 Copyright 69
Chapter 4 : Students' Reasoning in Chemistry Arguments and Designing Resources Using Constructive Alignment 93
4.1 Introduction to the research article presented in this chapter 93
4.2 Copyright 93
Chapter 5 : The role of language in students' justifications of chemical phenomena 110
5.1 Introduction to the research article presented in this chapter 110
5.2 Copyright 111
Chapter 6 : "I am working 24/7, but I can't translate that to you": The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language backgrounds 134
6.1 Introduction to the research article presented in this chapter 134
6.2 References for Introduction 135
6.3 Copyright 135
Chapter 7 : The experiences of chemistry researchers from diverse language backgrounds 150
7.1 Introduction to the research presented in this chapter 150
7.2 Introduction 151
7.2.1 English is the language of science, but not all scientists 151
7.2.2 Generative artificial intelligence (GenAI) to support linguistic equity 154
7.3 Goals and research questions 155
7.4 Theoretical Framework: Learning, communicating, and doing science 156
7.5 Methods 157
7.5.1 Questionnaire development 157
7.5.2 Overview of the questionnaire 157
7.5.3 Validity evidence 158
7.5.4 Content validity evidence 159
7.5.5 Response process validity evidence 160
7.5.6 Research Ethics Board approval 161
7.5.7 Pilot study 161
7.5.8 Data collection 162
7.5.9 Qualitative data analysis 162
7.5.10 Quantitative data analysis 163
7.6 Results and discussion 165
7.6.1 Research question 1: How does the need to use English in science impact their research experiences and perceived productivity? 165

7.6.2 Research question 2: What resources and strategies do they use to navigate language barriers during their research?....167
7.6.3 Research question 3: In what ways are participants using generative artificial intelligence, if at all?....171
7.6.3.1 Communicating in chemistry....172
7.6.3.2 Learning in chemistry....174
7.6.3.3 Doing chemistry research....176
7.6.3.4 Concerns and reasons for not using GenAI....179
7.7 Conclusions, implications, and future work....182
7.8 Limitations....183
7.9 References for Chapter 7....186
Chapter 8 : Response process validity evidence in chemistry education research....198
8.1 Introduction to the research article presented in this chapter....198
8.2 References for Introduction....198
8.3 Copyright....199
Chapter 9 : Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities....211
9.1 Introduction to the research article presented in this chapter....211
9.2 Copyright....212
Chapter 10 : Analysis of financial challenges faced by graduate students in Canada....223
10.1 Introduction to the research article presented in this chapter....223
10.2 Copyright....223
Chapter 11 : Prioritizing Mentorship as Scientific Leaders....260
11.1 Introduction to the research article presented in this chapter....260
11.2 Copyright....260
Chapter 12 : Conclusions and Final Reflection....267
12.1 Overview of research significance and implications....267
12.2 Overview of future research directions....271
12.3 Final reflection....274
12.4 References for Chapter 12....276
Chapter 13 : Appendices....281
13.1 Appendix 1: Chapter 2 Supporting Information....281
13.2 Appendix 2: Chapter 5 Supporting Information....293
13.3 Appendix 3: Chapter 6 Supporting Information....334

13.4 Appendix 4: Chapter 7 Supporting Information.....355
13.5 Appendix 5: Chapter 8 Supporting Information.....383
13.6 Appendix 6: Chapter 9 Supporting Information.....411
13.7 Appendix 7: Chapter 10 Supporting Information.....424
13.8 Appendix 8: Chapter 11 Supporting Information.....444

List of Figures

Figure 1: Sources of validity evidence, including definitions and examples of methods to collect each type of evidence. Adapted from Lapierre (2019). 8
Figure 2: Differences in questionnaire responses between Eng+ chemists (n = 126) and Eng1st chemists (n = 99) to questionnaire sections focused on (1) difficulties with English; (2) challenges with research due to English; and (3) perceptions of English impact on research productivity. Dashed line = mean; Non-dashed line = median. 166
Figure 3: Differences in reported helpfulness of different resources between Eng+ and Eng1st respondents. *$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$ 168
Figure 4: Overview of codes from participants' responses to "How do you think others in thes scientific community could better support scientists who are learning English, if at all? Please try to be as detailed as possible." 169
Figure 5: Reported uses of GenAI tools for language support in research by both Eng1st (n = 99) and Eng+ (n = 126) participants. 172
Figure 6: Uses and benefits of GenAI tools for communicating in chemistry reported by participants. ..173
Figure 7: Uses and benefits of GenAI tools for learning in chemistry reported by participants. 175
Figure 8: Uses and benefits of GenAI tools for doing chemistry research reported by participants. 177
Figure 9: Limitations and concerns about GenAI tools for chemistry research reported by participants. 179
Figure 10: Reasons for not using GenAI tools in chemistry research reported by participants. 181

List of Tables

Table 1: Descriptions and examples of the modes of reasoning. From Deng and Flynn (2021). 17

Overview of Academic Contributions

Through my doctoral studies, I have been driven by my deep desire to help people learn and make learning in chemistry and science more equitable and inclusive.

Collectively, the research projects I have contributed to or led during my doctoral studies have contributed to multiple areas of chemistry education research and beyond. Our peer-reviewed manuscripts and book chapters have been cited over 160 times and viewed over 25,000 times. Our work has been published in field-specific journals (*Journal of Chemical Education*, *Chemistry Education Research and Practice*), as well as journals with broader readership (*International Journal of Science Education, ACS Central Science*). Our work on reasoning and argumentation in chemistry education has informed many new studies and initiatives, including investigations of students' reasoning in chemistry and science and the development of new teaching approaches and formative assessment tools to support the development of students' reasoning (*e.g.*, scaffolding, machine learning tools). Our work on language in chemistry education and research is amongst the first in chemistry and science education research, and it will serve as a basis for many future initiatives and studies on linguistic equity in chemistry and science. Our methods manuscript on response process validity evidence in CER has already been a valuable resource for the CER community and beyond, and it will continue to be a foundational text as our field continues to grow. My work on chemistry graduate student EDI communities, financial challenges of graduate students in Canada, and mentorship in chemistry have shaped and will continue to shape conversations about the importance of graduate student leadership and education in science.

I have received national and international awards for my academic contributions. Most notably: I am 2021 Vanier Scholar, 2022 CAS Future Leader, and was selected as the inaugural J.C. Tito Scaiano Graduate Student of the Year at uOttawa in 2021. Beyond my research, I have been involved in various national initiatives and committees related to enhancing equity, diversity, and inclusion (EDI) in Canadian chemistry and higher education, including the Canadian Society of Chemistry's Working for Inclusivity, Diversity, and Equity (WIDE) Committee and the Society of Teaching and Learning in Higher Education's (STLHE) EDI Task Force, and the Ottawa Science Policy Network (OPSN). I love

teaching and mentorship and was selected as the 2020 Teaching Assistant of the Year by uOttawa's Science Students' Association (SSA).

List of Contributions

Publications (*corresponding author)

1. ***Deng, J. M.;** Ahmed, S. E.‡; Awoonor-Williams, E.‡; Banerjee, P.‡; Barecka, M. H.‡; Bickerton, L. E.‡; Di Pietro, S. A.‡; Dorn, S. K.‡; Jablonka, K. M.‡; Laudadio, G.‡; Kreidt, E.‡; Mannochio-Russo, H.‡; Terra, J.‡; Wilkins, O. H.‡; Yerneni, S. S.‡; Yusuf, M.‡ Prioritizing Mentorship as Scientific Leaders. *ACS Cent. Sci.* **2024**, *10* (2), 209–213. doi.org/10.1021/acscentsci.3c00500
2. ***Deng, J.M.**; Lalani, Z.; McDermaid, L. A.; Szozda, A.R. Using Generative Artificial Intelligence in Chemistry Education Research: Prioritizing Ethical Use and Accessibility. Submitted. Available on ChemRxIV.
3. **Deng, J. M.;** Flynn, A. B. "I Am Working 24/7, but I Can't Translate That to You": The Barriers, Strategies, and Needed Supports Reported by Chemistry Trainees from English-as-an-Additional Language Backgrounds. *J. Chem. Educ*. **2023**, *100* (4), 1523–1536. doi.org/10.1021/acs.jchemed.2c01063.
4. Laframboise, S. J.; Bailey, T.; Dang, A.-T.; Rose, M.; Zhou, Z.; Berg, M. D.; Holland, S.; Abdul, S. A.; O'Connor, K.; El-Sahli, S.; Boucher, D. M.; Fairman, G.; **Deng, J.M.**; Shaw, K.; Noblett, N.; D'Addario, A.; Empey, M.; Sinclair, K. Analysis of Financial Challenges Faced by Graduate Students in Canada. *Biochem. Cell Biol.* **2023**, *101* (4), 326–360. doi.org/10.1139/bcb-2023-0021.
5. **Deng, J. M.;** Rahmani, M.; Flynn, A. B. The Role of Language in Students' Justifications of Scientific Phenomena. *Int. J. Sci. Educ.* **2022**, *44* (13), 2131–2151. doi.org/10.1080/09500693.2022.2114299.
6. ***Deng, J. M.;** McMunn, L. E.; Oakley, M. S.; Dang, H. T.; Rodriguez, R. S. Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities. *J. Chem. Educ*. **2022**, *99* (1), 373–382. doi.org/10.1021/acs.jchemed.1c00485.
7. **Deng, J. M.;** Streja, N.; Flynn, A. B. Response Process Validity Evidence in Chemistry Education Research. *J. Chem. Educ.* **2021**, *98* (12), 3656–3666. doi.org/10.1021/acs.jchemed.1c00749.
8. Carle, M. S.; **Deng, J. M.**; Huang, D.; Lapierre, K.; Mesnic, N.; Bodé, N. E.; Featherstone, R. B.; Ingram, Q.; O'Connor, E. K.; Roberge, S.; et al. Engaging Students as Co-Creators of Course Resources and Learning Experiences. *New Dir. Teach. Learn.* **2021**, *167*, 63–76. doi.org/10.1002/tl.20460
9. **Deng, J. M.;** Flynn, A. B. Reasoning, Granularity, and Comparisons in Students' Arguments on Two Organic Chemistry Items. *Chem. Educ. Res. Pract.* **2021**, *22*, 749–771. doi.org/10.1039/D0RP00320D
10. Bodé, N. E., **Deng, J. M.**, & Flynn, A. B. Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *J. Chem. Educ.* **2019**, *96*(6), 1068–1082. doi.org/10.1021/acs.jchemed.8b00719

Books and book chapters (‡equal contribution)

1. **Deng, J.M.‡**, Carle, M.S.‡, Flynn, A.B. "Supporting students' reasoning in chemistry through constructive alignment." In *Student Reasoning in Organic Chemistry*. 2022.
2. Mighty, J.; Braithwaite, A.; **Deng, J.M**.; Huang, S.; Leacock, B.; Lopes, V.; MacDonald, M.; Pérez, R.N.; Owusu, B.; Toulouse, P. "Misalignment, Negotiation, and Compromise: The Challenging Process of Pursuing Equity, Diversity, and Inclusion in Higher

Education" In *Taking Stock 2.0: Transforming Teaching and Learning in Higher Education*; 2022.

3. **Deng, J.M.**‡; Doherty, E.‡; Young, L.P. "Battling Spectres of Precarity: Teaching Graduate Students to Model and Market Professional Skills as Wellness Practice." In *Taking Stock 2.0: Transforming Teaching and Learning in Higher Education*; 2022.

Invited Presentations and Workshops

1. Deng, J.M. "Using evidence to foster a more equitable and inclusive chemistry community" (Oral), University of Michigan, Ann Arbor, Michigan, USA, Feb 2024.
2. Deng, J.M. "Using evidence to foster a more equitable and inclusive chemistry community" (Oral), University of Ottawa, Ottawa, Ontario, Canada, Nov 2023.
3. Deng, J.M.; Flynn, A.B. "'I am working 24/7, but I can't translate that to you': The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language background" (Oral), Reed College, Portland, Oregon, USA, Sept 2023.
4. Deng, J.M.; Flynn, A.B. "'I am working 24/7, but I can't translate that to you': The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language background" (Oral), Carleton Biology Seminar Series, Carleton University, Ottawa, Ontario, Canada, Jan 2023.
5. Deng, J.M.; Flynn A.B. "Language in Chemistry", Invited speaker for CHEM 550 course (Chemistry Education Research and Practice), University of Michigan, Ann Arbor, Michigan, USA, Nov 2022.
6. Deng, J.M.; Flynn, A.B. "The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language (Eng+) backgrounds" (Invited Seminar), Department of Chemistry, Cape Breton University, Sydney, Nova Scotia, Canada, Nov 2022.
7. Deng, J.M.; Streja, N.S.; Flynn, A.B. "Collecting response process validity evidence in chemistry education research" (Oral), Biennial Conference on Chemical Education (BCCE). West Lafayette, Indiana, USA, Aug 2022.
8. Deng, J.M.; Flynn, A.B. "The experiences of chemistry graduate students, post–doctoral fellows, and faculty from English-additional language (Eng+) backgrounds" (Oral) Biennial Conference on Chemical Education (BCCE). West Lafayette, Indiana, USA, Aug 2022.
9. Deng, J.M.; Flynn, A.B. "Characterizing students' argumentation in chemistry education research" (Oral), Atkinson Research Group Meeting, February 2022.
10. Deng, J.M. "Evidence-based practices to foster an equitable and inclusive chemistry community" (Oral), joint workshop session hosted by Canadians Working for Inclusivity in the Chemical Sciences (CWIC) Network and Chemical Institute of Canada. February 2022.
11. Deng, J.M. "Race and racism as an Asian American in STEM" (Oral), Leaders Overcoming Great Inequalities in Chemistry (LOGIC). August 2021.
12. Deng, J.M. "Catalysts for Change: Tackling Racism in Chemistry and Beyond" (Oral), University of Toronto, Canada, May 2021.
13. Deng, J.M. "Sustaining a culture of inclusion in chemistry through student diversity, equity, and inclusion learning communities" (Oral), University of Iowa, USA, March 2021.

14. Deng, J.M. “Embedding anti-racism throughout the chemical sciences” (Oral), Canadian Chemical Engineering Conference (CCEC). Ottawa, Ontario, Canada, October 2020.
15. Deng, J.M.; Flynn, A.B. “Characterizing students’ abilities to argue about ‘why’ on organic chemistry assessments” (Oral), American Chemical Society (ACS) National Meeting. San Diego, California, USA, August 2019.
16. Deng, J.M.; Flynn, A.B. “Characterizing students’ abilities to argue about ‘why’ on organic chemistry assessments” (Poster), American Chemical Society (ACS) National Meeting. San Diego, California, USA, August 2019.

Contributed Presentations (*received presentation award)

1. Deng, J.M.; Flynn, A.B. “The experiences of chemistry trainees learning English-as-an-additional language (Eng+)” (Oral), Chemistry Education Research and Practice, Gordon Research Seminar. Lewiston, Maine, USA, July 2023.
2. Deng, J.M.; Flynn, A.B. “The experiences of chemistry trainees learning English-as-an-additional language (Eng+)” (Poster), Chemistry Education Research and Practice, Gordon Research Conference. Lewiston, Maine, USA, July 2023.
3. Deng, J.M.; Flynn, A.B. “’I am working 24/7, but I can't translate that to you’: The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language background” (Oral), Canadian Society of Chemistry (CSC) National Meeting. Vancouver, British Columbia, Canada, June 2023.
4. Deng, J.M. “Beyond the model minority: equity and inclusion for Asians in STEM” (Oral), Canadian Society of Chemistry (CSC) National Meeting. Vancouver, British Columbia, Canada, June 2023.
5. Deng, J.M.; Flynn, A.B. “’I am working 24/7, but I can't translate that to you’: The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language background” (Oral), X-DBER. University of Nebraska-Lincoln (Virtual), Apr 2023.
6. Deng, J.M.; Flynn, A.B. “Focus groups with chemistry graduate students and faculty from English-additional language (Eng+) backgrounds: Barriers, strategies, and recommendations” (Oral) American Chemical Society (ACS) National Meeting. Chicago, Illinois, USA, August 2022.
7. Deng, J.M.; Flynn, A.B. “Students’ justifications of chemistry phenomena: The role of language proficiency” (Oral), American Chemical Society (ACS) National Meeting. Chicago, Illinois, USA, August 2022.
8. Deng, J.M.; Flynn, A.B. “The experiences of chemistry graduate students, post-doctoral fellows, and faculty from English-additional language (Eng+) backgrounds” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Calgary, Alberta, Canada, June 2022.
9. Deng, J.M.; Streja, N.S.; Flynn, A.B. “Collecting response process validity evidence in chemistry education research” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Calgary, Alberta, Canada, June 2022.
10. Deng, J.M.; Rahmani, M.; Flynn, A.B. “The role of language proficiency in students' justifications of chemical phenomena” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Calgary, Alberta, Canada, June 2022.

11. Deng, J.M.; Flynn, A.B. “The experiences of chemistry graduate students, post–doctoral fellows, and faculty from English-additional language (Eng+) backgrounds” (Oral), Ontario–Carleton Chemical Institute (OCCI) Day. Ottawa, Ontario, Canada, June 2022.
12. Deng, J.M.; Rahmani, M.; Flynn, A.B. “Language proficiency and students’ abilities to justify chemical phenomena” (Oral), Pacifichem. Honolulu, Hawaii, USA, December 2021. *Finalist for Student Research Competition.*
13. Deng, J.M.; Rahmani, M.; Flynn, A.B. “The role of language proficiency in students’ abilities to justify chemical phenomena” (Oral), International Union of Pure and Applied Chemistry (IUPAC) and Canadian Chemistry Conference and Exhibition (CCCE). Montreal, Québec, Canada, August 2021.
14. Deng, J.M.; Rahmani, M.; Flynn, A.B. “Students’ language backgrounds and how they reason about chemical phenomena” (Oral), Ottawa-Carleton Chemical Institute (OCCI) Day. Ottawa, Ontario, Canada, July 2021.
15. Deng, J.M. “English learners in chemistry education” (Group discussion), Chemistry Education Research Literature Club. August 2021.
16. Deng, J.M. “Analyzing reasoning in students’ written arguments and explanations: opportunities, challenges, and equity considerations” (Poster), Methods in Chemistry Education Research. June 2021.
17. Deng, J.M.; Flynn, A.B. “Reasoning, granularity, and comparison in students’ arguments on two organic chemistry items” (Poster), Royal Society of Chemistry Twitter Conference. March 2021.
18. *Deng, J.M.; Flynn, A.B. “Evaluating the impact of English language proficiency on students’ abilities to construct chemistry arguments” (Oral), Chemistry and Biochemistry Graduate Research Conference. Montreal, Québec, Canada, November 2020.
19. Deng, J.M.; Flynn, A.B. “Investigating the impact of English language proficiency on chemistry students’ arguments” (Oral), Chemistry and Synthetic Biology Symposium. Ottawa, Ontario, Canada, October 2020.
20. Deng, J.M.; Flynn, A.B. “Characterising students’ arguments in three dimensions: reasoning, granularity, and comparisons” (Poster), Methods in Chemistry Education Research. June 2020.
21. *Deng, J.M.; Flynn, A.B. “How students justify their claims on organic chemistry assessments” (Oral), Ontario–Carleton Chemical Institute (OCCI) Day. Ottawa, Ontario, Canada, June 2020.
22. *Deng, J.M. “The invisible minority: strategies for supporting Asians and Asian Canadians in STEM” (Poster), Leaders Overcoming Gender Inequality in Chemistry (LOGIC). May 2020.
23. Deng, J.M.; Flynn, A.B. “A unit–based method for characterizing students’ scientific arguments on chemistry assessments” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Winnipeg, Manitoba, Canada, May 2020. Cancelled due to COVID-19.
24. Deng, J.M. “The invisible minority: strategies for supporting Asians and Asian Canadians in STEM” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Winnipeg, Manitoba, Canada, May 2020. Cancelled due to COVID-19.
25. Streja, N.; Deng, J.M.; Flynn, A.B. “Investigating How Students Discuss Energy by Analyzing Their Constructed Reaction Coordinate Diagrams” (Oral), Southern Ontario Undergraduate Student Chemistry Conference. 2020.

26. Streja, N.; Deng, J.M.; Flynn, A.B. “Investigating How Students Discuss Energy by Analyzing Their Constructed Reaction Coordinate Diagrams” (Oral), University of Ottawa Honours Symposium. 2020.
27. Deng, J.M.; Bodé, N.E.; Flynn, A.B. “Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms” (Poster), Royal Society of Chemistry Twitter Conference. March 2020.
28. *Deng, J.M.; Flynn, A.B. “Characterizing students’ abilities to construct scientific arguments in organic chemistry” (Poster), Chemistry and Synthetic Biology Symposium. Ottawa, Ontario, Canada, October 2019.
29. Deng, J.M.; Flynn, A.B. “Characterizing students’ abilities to reason about ‘why’ on organic chemistry assessments” (Oral), Canadian Chemistry Conference and Exhibition (CCCE). Quebec City, Quebec, Canada, June 2019.
30. *Deng, J.M.; Flynn, A.B. “Characterizing students’ abilities to reason about ‘why’ on organic chemistry assessments” (Oral), Ontario–Carleton Chemical Institute (OCCI) Day. Ottawa, Ontario, Canada, June 2019.
31. Deng, J.M.; Flynn, A.B. “Incorporating systems thinking into chemistry education through the United Nations’ Sustainable Development Goals” (Poster), Ottawa-Carleton University 2nd Annual Cross–Faculty Symposium on Science Teaching and Learning. Ottawa, Ontario, Canada, October 2018.

List of Abbreviations

Eng+	English-as-an-additional language
Eng1st	English-as-a-first language
EDI / DEI	Equity, diversity, and inclusion / Diversity, equity, and inclusion
CER	Chemistry education research
DBER	Discipline-based education research
KMb	Knowledge mobilization
ILO	Intended learning outcome
REB	Research Ethics Board
TCPS2	Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans
CGEDIC	Chemistry graduate equity, diversity, and inclusion communities
NSERC	Natural Science and Engineering Research Council
SSHRC	Social Science and Humanities Research Council
CIHR	Canadian Institutes of Health Research
STEMM	Science, technology, engineering, mathematics, and medicine
AAAS	American Association for the Advancement of Science
RSC	Royal Society of Chemistry
E1	Unimolecular elimination
E2	Bimolecular elimination
S_N1	Unimolecular nucleophilic substitution
S_N2	Bimolecular nucleophilic substitution
OC I	Organic Chemistry I
OC II	Organic Chemistry II

Chapter 1: Introduction

The Introduction of my thesis provides readers with an overview of the topics, ideas, problems, and frameworks that have guided the various research projects I have contributed to or led during my doctoral studies. The chapter begins with an overview of chemistry education research to supply readers with a general understanding of the field's history, goals, and approaches. The chapter then details topics, ideas, problems, and frameworks relevant to my doctoral research, with the relevant projects and chapters detailed at the end of each section. The chapter concludes with an overview of this thesis's structure and content.

1.1 Introduction to Chemistry Education Research (CER)

1.1.1 What is chemistry education research and why is it important?

We are living in a world facing many complex challenges. Solving these challenges will require approaches to science education that can effectively equip future leaders and citizens with core disciplinary knowledge and scientific skills (National Research Council, 2013; Organisation for Economic Cooperation and Development, 2006; Social Sciences and Humanities Research Council, 2018). Discipline-based education research (DBER) helps support these goals by helping: (1) understand how people learn the concepts, practices, and ways of thinking of science and engineering; (2) understand the nature and development of expertise in a discipline; (3) help to identify and measure appropriate learning objectives and instructional approaches that advance students toward those objectives; (4) contribute to the knowledge base in a way that can guide the translation of DBER findings to classroom practice; (5) and identify approaches to make science and engineering education broad and inclusive (National Research Council, 2012b).

DBER that can be defined as educational research grounded in the "priorities, worldview, knowledge, and practices" of the discipline of interest (National Research Council, 2012b).

Chemistry education research (CER) is a type of DBER that involves the systematic study of teaching and learning in chemistry, investigated using a variety of qualitative and quantitative methods (M. M. Cooper & Stowe, 2018). CER (and other forms of DBER) are unique from other educational research fields (*e.g.*, scholarship of teaching and learning, educational psychology, cognitive science) in that researchers in CER must not only have knowledge of and be skilled in broader educational research theory, principles, and methods, but also topics, knowledge, issues, priorities, practices, *etc*. unique to the discipline of chemistry (National Research Council, 2012b). For instance, consider a hypothetical study focused on the learning and research experiences of chemistry graduate students. The perspectives and insights of chemistry education researchers are essential in this type of study, as knowledge of the chemical enterprise, the career and learning progression of chemistry trainees, *etc.* provides a valuable vantage point for deeper understanding and inquiry (M. M. Cooper & Stowe, 2018).

There are three broad categories of CER (Taber, 2013). Inherent CER focuses on topics that "arise from the specific nature of teaching and learning chemistry as a curriculum subject". An example of inherent CER might be exploring students' interpretations of the electron-pushing formalism, a symbolism unique and inherent to chemistry (A. B. Flynn & Featherstone, 2017; Galloway et al., 2017, 2019; Youssef, 2023). Embedded CER focuses on topics that "arise from general issues in teaching and learning, but have been conceptualized within the specific context of teaching and learning chemistry." An example of embedded CER might be exploring students' self-efficacy beliefs (a general topic) within the specific context of learning chemistry (Lapierre, 2019). Collateral CER focuses on topics that "arise from, and are conceptualized as, general issues in teaching and learning, and chemistry teaching/learning simply provides a convenient context for data collection." An example of collateral CER might be exploring the effects of a self-regulated learning and goal-setting intervention designed to be usable for all students, with students in a chemistry course simply providing a convenient population for data collection (A. Flynn et al., 2020; O'Connor et al., 2021).

1.1.2 How is chemistry education research conducted?

In the following sections, I provide an overview of the role of research questions, theoretical frameworks, qualitative methods and quantitative methods (including data collection and analyses), validity and reliability, and knowledge mobilization in chemistry education research.

1.1.2.1 Research questions in chemistry education research

Research questions serve as the foundation for any investigation or study, articulating the knowledge that researchers seek through their research, and defining the scope and direction of the study, including data collection, data analysis, and knowledge mobilization. Preparing clear and effective research questions is a skill that can be refined with practice, resources, and support. One framework we have used in the Flynn group for preparing research questions prompts researchers to consider the following questions: (1) Is the question worth asking? (2) Is it feasible? (3) Who or what will be studied? (4) How will it be investigated? (5) What is the potential "take home" message of the investigation? (Bunce, 2008).

1.1.2.2 Theoretical frameworks in chemistry education research

A theoretical framework serves as a guiding lens for chemistry education research studies. A theoretical framework is a system of ideas, aims, goals, and assumptions about knowledge; about how research should be carried out; and about how research should be reported that influences what kind of experiments can be carried out and the type of data that result from these experiments (Bodner & Orgill, 2007; Creswell, 2012). Theoretical frameworks help researchers provide a theoretical basis for assumptions, connecting the researcher to existing research or theory, and helping researchers define the limits and boundaries of their research.

Choosing a theoretical framework involves a deliberate process aimed at aligning the research objectives with an appropriate conceptual and theoretical lens. Researchers need to consider the research question and objectives, identifying the key concepts and variables involved. They may also review existing literature within the field to identify established frameworks that are relevant

to their research topic. In this process, researchers assess the strengths and limitations of each potential framework in relation to the research context, considering factors such as relevance, explanatory power, and empirical support. Additionally, researchers should evaluate their own theoretical orientation and preferences, as well as the feasibility of operationalizing the chosen framework within the research design. Ultimately, the chosen theoretical framework should offer a coherent and comprehensive perspective that guides the research process and facilitates meaningful interpretation of findings.

1.1.2.3 Ethics in chemistry education research

Chemistry education research often involves human participants. In Canada, the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) provides ethical guidelines research involving human participant, guided by three core principles: Respect for Persons, Concern for Welfare, and Justice (Government of Canada Tri-Agencies, 2022). It states that participants are volunteers who have the right to informed consent, to choose to withdraw from studies at any point, and to not be harmed from studies. In all research involving human participants, researchers cannot conduct their research unless first reviewed and approved by research ethics boards (REBs). All my research received approval from the University of Ottawa's Research Ethics Board.

1.1.2.4 Qualitative and quantitative methods in chemistry education research

Quantitative methods in chemistry education research involve collecting, analyzing, and interpretating data to quantify and measure phenomena, relationships, and outcomes in chemistry education (Creswell, 2012). Researchers often aim to develop a *breadth* of understanding. Common goals of quantitative studies include: calculating descriptive statistics (*e.g.*, mean, median, standard deviation), and determining whether quantified phenomena are statistically generalizable and testing hypotheses using statistical analysis. Because statistical power depends on sample size, larger sample sizes are often required for quantitative methods. In the wet lab,

quantitative methods could be considered akin to calculating a synthetic yield or identifying reaction rates.

Qualitative methods in chemistry education research involve collecting, analyzing, and interpreting data in a descriptive, non-numerical form (Creswell, 2012). These methods are used to understand and explore chemistry education phenomena, explore individuals' experiences, perspectives, and behaviors within chemistry education, and uncover underlying meanings and patterns. Researchers often aim to develop a depth of understanding and often collect data from smaller sample sizes due to the time and resources required for qualitative methods. Some common qualitative methods used in CER include interviews and focus groups. In the wet lab, qualitative methods are like mechanistic studies in which a chemist seeks to understand how and why the transformation occurs mechanistically.

A researcher's decision to use to qualitative or quantitative methods will depend on the research question and goal. For example, a researcher may be in the following question: what is the effect of a learning intervention on student learning? To address this question, a quantitative study could be designed in which the exam scores of students who receive the intervention versus those a control group are compared using statistical analysis. Through this, the researcher would develop an understanding of the effect of the intervention of student learning, but not about the underlying mechanism. To understand the underlying mechanism, a complementary qualitative study could be designed, such as interviews with students to probe their perspectives on what aspects of the intervention were (un)helpful.

The choice of methods may also depend on practical factors, such as time, resources, and available sample size. For example, the researcher may have only a sample size of 10 participants, limiting the statistical power and confidence of any claims drawn from any quantitative methods, and making a qualitative study more feasible.

1.1.2.5 Instruments in chemistry education research

Instruments are invaluable tools of measurement for researchers used to directly and indirectly observe otherwise inaccessible phenomena (Campbell & Nehm, 2013; Opfer et al., 2012). For example, chemists regularly use instruments (*e.g.*, spectrometers, high pressure liquid chromatographer) to measure physical and chemical entities and phenomena, with measurements providing the user with evidence related to the entities and phenomena being measured. Similarly, chemistry education research instruments—including assessments, tests, surveys, and observation protocols—often seek to measure variables within humans (respondents) directly and indirectly (Bollen, 2002; Opfer et al., 2012). Among the goals of discipline-based education research (DBER), one is to "help identify and measure appropriate learning objectives and instructional approaches that advance students toward those objectives" (National Research Council, 2012b). As DBER has grown, so too have the number of instruments designed to investigate different variables of interest (Atkinson et al., 2020; Denofrio et al., 2007; Sirum & Humburg, 2011; Smith et al., 2008, 2013; Wren & Barbera, 2013).

Instruments ideally generate valid data—which refers to the extent to which a test or assessment measures what it claims to measure—and reliable data—which refers to the consistency or stability of a test or measurement tool (American Psychological Association et al., 2014; Cronbach, 1971; Kane, 2013; Messick, 1995). The following sections provide an overview of validity and reliability in both qualitative and quantitative research.

1.1.2.5.1 Validity in quantitative research

Validity refers to the degree to which an instrument is measuring what the instrument is intended to measure (American Psychological Association et al., 2014; Cronbach, 1971; Kane, 2013; Messick, 1995). Collecting validity evidence is imperative in instrument development—without evidence for validity of the data obtained from a particular instrument in specific research context, there is uncertainty about what the instrument is actually measuring (American Psychological Association et al., 2014).

In some cases, instruments may not yet exist for a particular goal, requiring instruments to be created and supported with appropriate validity evidence to determine whether they measure what they intend to measure. The effectiveness and suitability of an instrument in chemistry education research can vary depending on the goals and context of a particular evaluation. An instrument designed for a particular purpose in one context may need to be re-evaluated, re-configured, or re-designed to provide meaningful measurements in pursuit of new research goals and/or within a new context. For example, an instrument designed to measure a particular variable within a university context may produce less valid data when used in a high school context.

Researchers can evaluate validity by collecting validity evidence, including instrument content, response process, internal structure, relationships with other variables, and consequences of instrument use (Figure 1) (American Psychological Association et al., 2014; Messick, 1995).

Instrument content validity ensures that the test or assessment adequately covers all relevant aspects of what it's supposed to measure. For example, a driving test that includes questions about traffic signs, road rules, and vehicle handling accurately represents the content necessary to evaluate someone's ability to drive safely.

Response process validity examines how participants understand and respond to the test items. For instance, after taking a reading comprehension test, students may be interviewed to understand how they interpreted the questions and what strategies they used to find answers.

Internal structure assesses how well the items within the test relate to each other and whether they consistently measure the intended construct. An example would be analyzing the results of a personality test to see if the questions group together in meaningful ways, reflecting different personality traits like extroversion or conscientiousness.

Relationships with other variables determines whether the scores on the test are related to other variables in ways that make sense. This could involve correlating scores on a job performance evaluation test with actual job performance ratings to see if there's a meaningful relationship between the two.

Consequences of instrument use considers the potential impact or consequences of using the test in real-world situations. For instance, evaluating whether using a standardized test for college admissions results in fair and equitable outcomes for all applicants, or if it disproportionately advantages certain groups.

Each type of evidence can be collected using specific and established research methods; readers interested in detailed definitions of each of these types of validity evidence and their collection methods are referred to the *Standards for Educational and Psychological Testing* (American Psychological Association et al., 2014).

During my doctoral studies, we collected evidence for response process and content validity for research described in Chapters 5 and 7. We also published a methods manuscript on collecting response process validity evidence in chemistry education research (Chapter 8) (Deng et al., 2021).

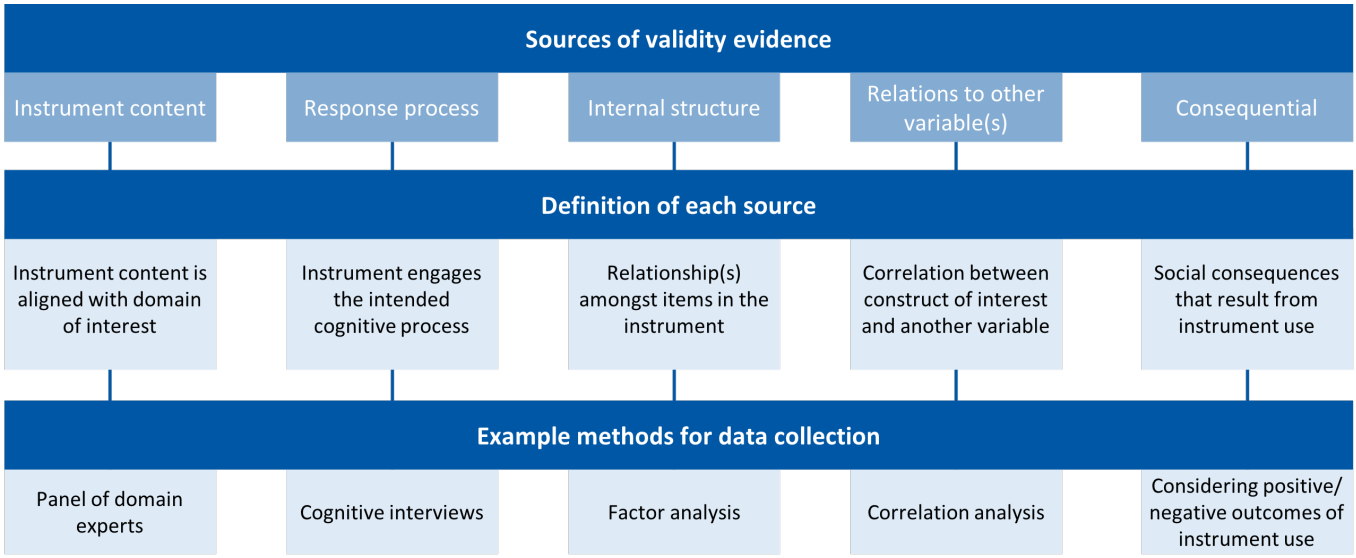


Figure 1: Sources of validity evidence, including definitions and examples of methods to collect each type of evidence. Adapted from Lapierre (2019).

1.1.2.5.2 Validity in qualitative research

Research is shaped by the perspectives and positionalities held by researchers (Curran & Randall, 2020; Darwin Holmes, 2020). In qualitative research, in which the goal is often to explore or gather a deeper understanding of a phenomenon, the concept of validity can be considered as a property of both the data and researchers' analyses of the data. For example, one common form of data in qualitative research is an audio recording from an interview. The recording itself could be considered more or less valid depending on whether these data produced from interviews align with the research goals and/or offer valid insights into interviewees thoughts, feelings, and perspectives. In this case, validity is a property of the data.

Once data are collected, researchers will often analyze these data for overarching patterns and themes, guided by the research goals. Here, validity is also relevant, but not in terms of the data generated but in terms of researchers' interpretations of the data. Researchers must take care in ensuring that their interpretations of the data are, as much as possible, valid interpretations of the data. Researchers can provide evidence and reasoning for their claims about the data in several ways, including collecting rich data, triangulating data, searching for disconfirming evidence, and member-checking (Creswell, J.W., & Miller, 2000; Hesse-Biber, 2010; Maxwell, 2012).

1.1.2.5.3 Reliability in quantitative research

Reliability in quantitative research refers to the consistency and stability of measurements or data collection techniques. Reliability reflects the extent to which the same results would be obtained if the research were repeated under identical conditions. For example, an NMR spectrometer produces reliable data if it can generate the same spectra for a compound, regardless of who uses the NMR or when or where the NMR is used. A CER instrument (*e.g.*, a test or interview protocol) generates reliable data if it can generate consistent data, regardless of who administers the questionnaire or the testing conditions. Ensuring reliability is crucial—if data are unreliable, it becomes challenging to draw accurate conclusions or make informed decisions based on the research outcomes.

Evaluating reliability in quantitative research typically involves various methods such as test-retest reliability and internal consistency reliability (American Psychological Association et al., 2014; Arjoon et al., 2013). Test-retest reliability assesses consistency by comparing results obtained from the same participants at different points in time. Internal consistency reliability evaluates the extent to which different items in a measurement tool consistently measure the same construct.

1.1.2.5.4 Reliability in qualitative research

A common concern during analysis of qualitative data is that one researcher's interpretations of data (*e.g.*, an interview transcript) will differ from another researcher's interpretations of the same data. In this case, reliability is in question, and the most common way to account for this potential variability (and correct for it) is inter-rater reliability (Garrison et al., 2006; Hallgren, 2012).

Inter-rater reliability begins with the two (or more) researchers independently coding the same set of data. Depending on the study, goals, and resources (time) available to researchers, this may involve both researchers coding the entire dataset individually, one researcher coding a subset of the data, *etc*. Once both researchers have analyzed the data, they can then use established mathematical approaches to calculate the level of agreement between their analyses. Common measures of agreement include percent agreement, Cohen's κ, and Krippendorff's α (Garrison et al., 2006; Hallgren, 2012). Each measure has its respective pros and cons—for example, κ can only be used to calculate agreement between two researchers and does not account for missing data, but κ *does* account for chance agreement, which percent agreement does not. Each measure also comes with different thresholds that define the level of agreement between researchers, including at what level agreement is deemed to be "acceptable". For example, for Krippendorff's α, <0.67 = poor agreement; $0.67 – 0.79$ = moderate agreement; > 0.80 = satisfactory agreement (Krippendorff, 1970).

Generally, if the researchers do not achieve a satisfactory level of agreement as defined by the chosen measure of inter-reliability, they will discuss disagreements and revise their coding approach accordingly. Often, a guide for how the researchers analyzed the data (called a codebook) will be generated and refined through these discussions. The researchers will then repeat the inter-rater reliability process until a satisfactory level of agreement is achieved.

Evaluating and maximizing reliability in qualitative research may not always be relevant, and will depend on whether the researchers value a positivist or postpositivist approach in their work. Positivists emphasize the importance of reliability as achieving consistent and replicable results, often through standardized methods and measurements, such as inter-rater reliability. In contrast, postpositivists recognize the complexity of human experiences and acknowledge that complete objectivity is unattainable. Postpositivists acknowledge the researcher's influence on the research process and actively engage in critical self-awareness to mitigate bias or centre certain perspectives and viewpoints that may generate more fruitful and valid data (Robson, 2002).

For example, consider a hypothetical study on the learning experiences of Indigenous students. Postpositivists argue that the identities and positionalities of the researchers (*e.g.*, whether they are Indigenous or not) can and will shape the research process, including the questions that are posed, the methods used, and how data analysis is conducted. For example, Indigenous members of the research team could bring unique perspectives rooted in their cultural identities, histories, and lived experiences that enrich the research process and potentially foster deeper understanding about Indigenous learners' experiences. In contrast, non-Indigenous researchers bring an outsider perspective, potentially laden with biases and power dynamics that could inhibit the research process and generate very different data and conclusions about learners. In these cases, reliability between the perspectives of the Indigenous and non-Indigenous researchers may not be important (and could even be harmful and inappropriate) if the research goal is to generate a valid and detailed understanding of Indigenous learners' experiences.

1.1.3 Knowledge mobilization

Knowledge mobilization (KMb) is the process of translating research findings, innovations, and expertise from research into practical applications and initiatives to benefit stakeholders (Graham et al., 2006; Social Sciences and Humanities Research Council, 2019). Knowledge mobilization is an essential part of the research process—it bridges the gap between research and real-world impact, ensuring that valuable insights generated through research are effectively utilized to address societal challenges or improve outcomes. KMb can include traditional modes of research dissemination, such as research publications and presentations. However, it also (and ideally) includes various other strategies to transform research into usable knowledge, such as creating workshops, creating resources (*e.g.*, digital media), and establishing partnerships, communities, *etc*.

Research and experts on KMb suggest that KMb efforts are most effective when they are interactive, two-way "exchange" processes—that is, they give equal importance to what researchers can learn from other stakeholders (A. Cooper, 2016; Lavis et al., 2003). KMb is also often an iterative and dynamic process—for example, researchers may host a workshop to mobilize knowledge, collect data outcomes and impacts from the workshop (participants' perspectives, changes in behaviour, *etc*.), and then use these data to refine the workshop.

KMb is particularly important in chemistry education research because the findings and implications of chemistry education research can often have a direct impact on the people with whom and places in which research was conducted. In our research on the experiences of chemistry trainees learning English-as-an-additional language (Eng+), we knew we needed to do more than publish a research article in a discipline-specific journal if we wanted to have an impact on the broader chemistry research community. We engaged in various strategic activities to mobilize our research findings into tangible recommendations and strategies, including: hosting seminars and workshops with faculty and students, sharing our work through news

outlets, and creating infographics and posters about our research (see List of Contributions, pages xiv–xviii).

1.1.4 Conclusion

All chemistry education research begins with research questions. These questions are used to guide all parts of the study, including the theoretical framework(s), data collection, data analyses, and knowledge mobilization. Methods in CER can be categorized as either quantitative or qualitative, and in both, validity (accuracy) and reliability (precision) are important for maximizing the quality and rigour of the data collected and analysis of that data. Once research is complete and researchers have identified the main findings and implications, knowledge mobilization activities can be conducted, which includes translating research findings, innovations, and expertise from research into practical applications and initiatives to benefit stakeholders.

During my doctoral studies, we employed a mix of quantitative and qualitative methods. We also employed mixed methods, which combine elements of qualitative and quantitative research methods within a single study or research project (*e.g.*, gathering both quantitative and qualitative data, conducting quantitative and qualitative analysis on the same dataset) (Wisdom & Creswell, 2013). We also collected evidence for different types of validity and reliability. We also employed a variety of theoretical frameworks, aligned with the different project goals and topics. The following sections will provide an overview of the topics, ideas, problems, and frameworks relevant to my doctoral research, with the relevant projects and chapters detailed at the end of each section.

1.2 Argumentation in chemistry education

As global issues such as the climate change and the COVID-19 pandemic continue to demonstrate, successfully navigating the complexity of present and future challenges facing our

world requires citizens and leaders capable of making and justifying decisions using scientific evidence (Social Sciences and Humanities Research Council, 2018; United Nations, 2015).

Educators, researchers, and organizations across the world have recognized the importance equipping students with the key scientific practice of scientific argumentation (European Union, 2006; National Research Council, 2012a; Organisation for Economic Cooperation and Development, 2006; Social Sciences and Humanities Research Council, 2018). These leaders have called for a scientific education that teaches learners not only about core disciplinary content knowledge, but also how to justify their claims and decisions with scientific evidence and reasoning. Aligned with this goal, many chemistry educators have worked to incorporate opportunities for students to develop and demonstrate their argumentation skills in chemistry courses and assessments (M. M. Cooper & Klymkowsky, 2013; Talanquer & Pollard, 2010), and chemistry education researchers have studied various dimensions and impacts of these activities (Becker et al., 2016; Bodé et al., 2019; M. M. Cooper et al., 2016; Deng & Flynn, 2021; Moon et al., 2016, 2019; Osborne & Patterson, 2011; Russ et al., 2008; Watts et al., 2020; M. L. Weinrich & Talanquer, 2016).

At its core, an argument is composed of three components: evidence, reasoning, and a claim (Toulmin, 1958). In an argument, one considers and presents evidence, reasoning with this evidence to justify a claim. Arguments therefore provide an opportunity for researchers and educators to investigate how individuals reason.

Chemistry education researchers have sought to understand students' reasoning through analyses of students' arguments about chemical phenomena disciplines (Becker et al., 2016; Bodé et al., 2019; M. M. Cooper et al., 2016; Deng & Flynn, 2021; Moon et al., 2016, 2019; Osborne & Patterson, 2011; Russ et al., 2008; Watts et al., 2020; M. L. Weinrich & Talanquer, 2016). In the following section, I provide an overview of different reasoning frameworks employed in chemistry education research, concluding with a modes of reasoning framework that has been employed in parts of my doctoral research.

1.2.1 Reasoning in chemistry education

Reasoning is the mental process of making sense of information, evaluating situations, and drawing conclusions or making decisions based on that information (Sternberg, 1986). Reasoning is a fundamental cognitive skill that humans use in problem-solving, decision-making, and understanding the world around them, and can be abstract, invisible, multi-dimensional, and dynamic (Krell et al., 2022; Talanquer, 2018b; Ward et al., 1990).

Reasoning has been conceptualized and analyzed through a variety of different frameworks in the learning and cognitive sciences. There is type I and II reasoning, in which reasoning is divided into intuitive, automatic thinking based on heuristics and past experiences (type I), while reasoning that involves deliberate, analytical thinking relying on logical reasoning and conscious evaluation of evidence (type II) (Evans, 2003; Sloman, 1996). Teleological reasoning involves explaining phenomena or events by their intended purpose or end goal, often attributing them to design or intentionality (Kelemen & Rosset, 2009). Abstractedness and abstraction in reasoning refer to the cognitive ability to detach from specific details and simplify complex information into generalized concepts or principles for more efficient problem-solving and analysis (Schwartz, 1995). Rules-based reasoning relies on predefined logical rules, case-based reasoning on past experiences, and model-based reasoning on mathematical or conceptual models to make decisions or solve problems (Kolodner, 1992; Nersessian, 1999).

In chemistry, the ability to reason is crucial. Incorporating student reasoning allows educators to evaluate students' abilities to think like scientists rather than simply evaluating their ability to store content taught in a course (Berland & Reiser, 2009). Each of the abovementioned frameworks have been applied to various studies in CER: type I and II reasoning (Maeyer & Talanquer, 2013; McClary & Talanquer, 2011; Talanquer, 2007, 2017); teleological reasoning (Abrams & Southerland, 2010; I Caspari et al., 2018; Ira Caspari et al., 2018; DeCocq & Bhattacharyya, 2019; Trommler et al., 2018); abstractedness and abstraction (Sevian et al., 2015;

Melissa L. Weinrich & Sevian, 2017); and rules-, case-, and model-based reasoning (DeCocq & Bhattacharyya, 2019; Kraft et al., 2010; Windschitl et al., 2008).

One reasoning framework in CER focuses on causal reasoning, which involves one's ability to identify and describe cause-and-effect relationships (Bodé et al., 2019; M. M. Cooper et al., 2016; Crandell et al., 2019, 2020; Deng et al., 2022; Deng & Flynn, 2021; Russ et al., 2008; Sevian & Talanquer, 2014). The ability of students to engage in causal reasoning is an important skill to develop as it can help prepare them for practical settings where a "correct answer" may not always be clear. For example, rather than relying on rules or memory to solve static problems, students who can employ causal reasoning strategies can apply their reasoning to a variety of situations and questions. Such skills are particularly pertinent in a world fraught with misinformation about scientific topics like vaccine mandates and climate change (Howell & Brossard, 2021; Osborne et al., 2013; Sadler et al., 2007; Sharon & Baram-Tsabari, 2020). By delving deeper into the "how" and "why", students can develop scientific thinking skills that rely on evidence to justify inferences, empowering them to question conventional "facts" and make well-informed decisions.

In Chapters 2–4, we employed a causal reasoning framework that conceptualizes reasoning as one of four modes (Table 1): descriptive, relational, linear causal, and multi-component causal. We used this framework to characterize students' reasoning demonstrated in their written arguments about chemical phenomena. This framework was first described in philosophy of science literature before being adapted for researchers and educators in chemistry education (Russ et al., 2008; Sevian & Talanquer, 2014). In this framework, **descriptive** reasoning is the most basic mode of reasoning. In this type of reasoning, claims are supported solely by stating rules or describing features and/or the properties of entities (e.g., the reactants, products, etc.), and no connections are established. **Relational** reasoning is an intermediate mode of reasoning in which relationships between properties of entities and their activities are discussed in a correlative fashion (i.e., a cause–effect relationship is not established). **Linear**

causal reasoning is an advanced mode of reasoning that establishes clear cause-and-effect relationships between the relevant properties of the entities and their activities. **Multi-component causal** reasoning is the most advanced mode of reasoning, with multiple cause-and-effect relationships between the relevant properties of the entities and their activities to support the claim. Though causal reasoning is the most sophisticated form of reasoning in this framework, the "best" or most "correct" form of reasoning will depend on the specific context or goal of the task/prompt.

Table 1: Descriptions and examples of the modes of reasoning. From Deng and Flynn (2021).

Mode of reasoning	Description	Example
Descriptive	Claims are supported solely by describing features and/or the properties of entities (e.g., the reactants, products, etc.). No connections are established.	NaH is the strongest base.
Relational	Relationships between properties of the reaction materials and their activities are discussed in a correlative fashion (i.e., a cause–effect relationship is not established)	NaH will drive the equilibrium towards products because its conjugate acid has a higher pK_a value (35) than the alkyne (25).
Linear causal	All of the features of a relational argument are present and are accompanied by cause-and-effect relationships between the relevant properties of the entities and their activities	NaH will drive the equilibrium towards products because its conjugate acid (pK_a value 35) is weaker/more stable than the starting alkyne (pK_a value 25), according to the relevant pK_a values. NaOH and NH_3 are not strong enough bases to drive the equilibrium to the product side, as both of their conjugate acids are stronger than the starting alkyne (pK_a values of 15.7 and 9, respectively).
Multi-component causal	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct interactions of several components	The hydride is the only base of the choices that is sufficiently strong to drive the equilibrium to products. The hydride base is smaller and less electronegative than carbon, making it less able to stabilize a negative charge than the alkyne anion (conjugate base). The oxygen atom in NaOH is more electronegative than the carbon atom in the conjugate base, making it better able to stabilize the negative charge. Although atomic

size (larger carbon atom) and hybridization suggest that the carbon atom in the conjugate base could be more stable, the pK_a values of the acid and conjugate acids support the conclusion that NaOH is the weaker base, since the stronger the acid, the weaker its conjugate base (pK_a value of H_2O = 15.7; pK_a of alkyne = 25) NH_3 is similarly too weak to drive the equilibrium to the product side, with the nitrogen atom being neutral in addition to being more electronegative than the carbon anion.

1.2.2 Granularity in science education

Arguments can also be constructed at different levels of granularity, also known as scales, levels nested hierarchies, emergence, and bottom-out reasoning (Darden, 2002; Luisi, 2002; Southard et al., 2017; Talanquer, 2018c; van Mil et al., 2013). For example, a justification for why aspirin acts as an acid in water can focus on pH and pK_a data (a phenomenological level of granularity) or instead how aspirin's carboxylic acid moiety is resonance-stabilized when deprotonated (underlying level that include discussion of structural, electronic, and energetic factors) (Talanquer, 2018a). Phenomena can be explained at different levels of granularity, from larger macroscopic levels (global levels and beyond) to smaller submicroscopic levels (atomic levels and below) (Darden, 2002; Machamer et al., 2000). The specific levels of interest will depend on the context and task. In Chapter 3, we characterized students' arguments in terms of four levels of granularity relevant to the prompts that were asked. The four levels of granularity were: phenomenological, energetic, structural, and electronic (Deng & Flynn, 2021).

The phenomenological level captures descriptions of chemical phenomena that arise from the interactions of molecules and atoms and their structural, electronic, and energetic properties. For example, within a given context, the favoured direction of a chemical equilibrium may be a phenomenon to be explained. The interplay of structural, electronic, and energetic properties/interactions of molecules and atoms can be used to determine and justify the direction of an equilibrium. Depending on the task, arguments may also be focused on other

phenomenological factors that can be equally valid; for example, pKa data could be used to determine the direction of an acid–base equilibrium.

The structural level captures descriptions of structural features of molecules and atoms. For example, in an acid–base equilibrium context, a student's description of the relative stability of two basic atoms would be considered discussion at the structural level. In the context of an organic chemistry mechanism, a structural discussion might include identifying steric bulk around a particular reactive centre and connecting the steric interactions to the effects at the transition state (energetic level). The structural level itself contains subdivisions, such as cells, biomolecules, small molecules, molecular fragments, functional groups, and individual atoms.

The electronic level captures descriptions of electronic features of molecules and atoms. For example, electronegativity and partial charges could be used to explain reactivity at the electronic level. Other examples might compare formal charges and electron density on basic atoms to justify the direction of an equilibrium or discuss molecular orbitals to describe electronic features of molecules. The energetic level captures descriptions of the energetics of reactions, including thermodynamic and kinetic considerations. Descriptions at this level could include considering the relative stabilities of conjugate acids/bases to justify the direction of an equilibrium or justifying the plausibility of various reaction mechanisms based on activation energies.

In Chapters 3, we selected and employed these four levels of granularity based on the concepts and ideas identified in students' responses, the intended learning outcomes related to the questions we analyzed, as well as previous theoretical work related to chemistry students' reasoning (Luisi, 2002; Machamer et al., 2000; Southard et al., 2017; Talanquer, 2018c; van Mil et al., 2013). Different levels of granularity may be more relevant for other contexts, such as other content areas within chemistry or other disciplines (biology, physics, *etc*.). For example, chemical reactions and equilibria may be the phenomena to be explained in the chemical contexts investigated in this study and the highest level of granularity needed for these contexts, while in

molecular biology contexts, these phenomena may be the deepest level of granularity needed for an explanation (e.g., explaining why a substrate binds to an enzyme).

Argumentation and reasoning are essential skills that students need to develop to thrive in a rapidly changing world. However, opportunities to develop and demonstrate these skills must also be equitable and inclusive for all learners, or else we risk exacerbating existing disparities in education and society. The following section describes these existing disparities, discussing the importance of equity, diversity, and inclusion in science/chemistry, and situating our work on linguistic equity in chemistry education and research.

1.3 Linguistic equity in chemistry education and research

Enhancing and sustaining equity, diversity, and inclusion (EDI) in science is essential for ensuring that all members of our community have opportunities to contribute fully and that we do not lose out on expertise and perspectives that might otherwise strengthen scientific progress (National Research Council, 2012b; Sanford, 2020; Winfield et al., 2020). Equity refers to the "the removal of systemic barriers and biases enabling all individuals to have equal opportunity to access and benefit from [a] program" (Government of Canada, 2020); Diversity to "differences in race, color, place of origin, religion, immigrant and newcomer status, ethnic origin, ability, sex, sexual orientation, gender identity, gender expression and age" (Government of Canada, 2020); and Inclusion to "the practice of ensuring that all individuals are valued and respected for their contributions and are equally supported" (Government of Canada, 2020)

Diversity in STEM is unrepresentative of the diversity in the general population (National Science Foundation, 2016, 2017; Whitcomb & Singh, 2020). A 2017 survey by the United States' National Science Foundation (NSF) found that women; Black, Indigenous, and People of Colour (BIPOC); and individuals with disabilities are underrepresented in the STEM workforce relative to the proportions of these groups within the general population (National Science Foundation, 2017). In Canada, Indigenous students are poorly represented in STEM relative to the general

population (Etkin et al., 2016). A 2018 survey by the Royal Society of Chemistry in the United Kingdom found that students with disabilities in chemistry face unique challenges that do not arise in other subjects; women in chemistry face barriers to achieving leadership roles, scientific publishing, and awards and the percentage of ethnic minority chemists drops significantly with increasing career stage (The Royal Society of Chemistry, 2018).

Causes for the lack of diversity observed in STEM are complex, multi-faceted, and the focus of ongoing work on equity and inclusion (Posselt, 2020; Sanford, 2020; Tilghman et al., 2021; Winfield et al., 2020). One challenge in improving diversity is the absence of diversity itself—poor diversity contributes to a self-perpetuating cycle in which students of equity-deserving groups struggle to progress in STEM due to lack of role models (Beasley & Fischer, 2012; Breda et al., 2018; Cheryan et al., 2015; González-Pérez et al., 2020). A lack of role models has also been associated with imposter syndrome experienced by scientists from equity-deserving groups (Funk & Parker, 2018). Another challenge is that support for scientific endeavours and access to a formal science education has historically excluded women, BIPOC, LGBTQ2S+ individuals, individuals with disabilities, and other equity-deserving groups (Woolston, 2020). This history has perpetuated a system in which certain individuals in society have greater access to education across generations, while students from equity-deserving groups are often first-generation students who lack role models within their personal, academic, and professional lives (Kredell, 2017). Who participates in science and engineering is also dependent on the state of DEIR at large—challenges faced by equity-deserving groups in STEM are a microcosm of the challenges faced by these groups in our broader society (D. T. Flynn, 2016; Fuentes & Liou, 2008; V. E. Lee & Burkam, 2002). Broader historical and societal factors contribute to implicit biases and stereotypes that undermine equity-deserving groups in various facets of STEM (coursework, hiring, *etc*.) (Barrett & Bliss-Moreau, 2009; Isgro & Castaneda, 2015; Moss-Racusin & Molenda, 2015; O'Brian et al., 2015; Reuben & Sapienza, 2014; Sarsons, 2015; Yoder & Mattheis, 2016).

One of the guiding goals of DBER is to "[identify] approaches to make science and engineering education broad and inclusive", highlighting a global need to deepen our understanding of evidence-based strategies to help broaden participation in STEM (National Research Council, 2012b). As the central science and a gateway into STEM disciplines (Holme & Hutchison, 2018), chemistry and chemistry education provides a strategic setting for addressing this goal.

Chapters 5–7 detail projects during my thesis focused on the learning and research experiences of chemistry students from diverse language backgrounds. In addition to the importance of EDI-focused chemistry education research in making chemistry more diverse, inclusive, and equitable, there are two primary reasons for focusing on this research topic.

The first reason is statistical. Chemistry classrooms and research labs around the world are and becoming more linguistically and culturally diverse. There are approximately 2 million international students studying in North America, many of whom come from nations where English is not a primary language (e.g., China, India, South Korea, Saudi Arabia) (Crossman et al., 2022; Higher Education Strategy Associates, 2020; Institute of International Education, 2020; National Academies of Science Engineering and Medicine, 2018). In Canada, international students make up over a quarter of the total post-secondary student population (Higher Education Strategy Associates, 2020). At the graduate student level, 38% of science and engineering graduate students are temporary visa holders (National Science Foundation, 2020). Many laboratories and scientific events rely on English, even in countries like Germany, Sweden, France, Switzerland, and Israel (Ammon, 2011). A shared language like English facilitates communication and globalization of science, but it may also limit science as many students and researchers today learn English-as-an-additional language (Eng+) (Amano et al., 2016, 2023; Ramírez-Castañeda, 2020).

The second reason is personal. Research is affected by the identities and perspectives of the researchers conducting the study (Curran & Randall, 2020; Howe et al., 2022). As a second-

generation Chinese immigrant born in Canada, I consider English to be my first language but grew up in a Cantonese-speaking home. My interest in the role of language in chemistry learning and research was prompted (in part) by reflections on how racism and xenophobia has impacted my life and the lives of those around me— including my parents, Eng+ immigrants.

1.3.1 Linguistic equity and scientific argumentation

Students can develop scientific argumentation skills through classroom activities and assessments (e.g., asking students to justify why a phenomenon occurs) but the linguistic skills required to construct arguments may be a challenge for students (McNeill et al., 2006; McNeill & Berland, 2010; McNeill & Krajcik, 2008a, 2008b). Constructing arguments in science can be especially challenging for students learning English-as-an-additional language (Eng+) who are expected to learn, use, and communicate abstract scientific concepts in academic language while they are simultaneously learning English (Abedi et al., 2020; Afitska & Heaton, 2019; Buxton et al., 2014; Curtis & Millar, 1988). Students' challenges with language may also be exacerbated in disciplines like chemistry in which terms like "stability" and "resonate" have different meanings from their everyday definitions (Childs et al., 2015; Uhl Chamot, 1995).

Eng+ students face unique linguistic challenges with learning, communicating and doing science (O. Lee & Fradd, 1998). Knowing science means making meaning of scientific knowledge by connecting new information to previous knowledge. Language is foundational to learning, as it mediates how learners engage with scientific concepts and construct meaning. Learning new information can be a challenge for Eng+ students, as they may have learned prerequisite concepts in a language or culture different from English, making it more challenging to identify and establish coherent connections between new information and their previous knowledge. Doing science involves scientific practices, such as engaging in inquiry, proposing arguments and explanations, and interpreting and identifying evidence to make sense of the world (O. Lee et al., 2013; O. Lee & Fradd, 1998). Many of these practices require complex language

functions and academic traditions that require sufficient levels of language proficiency within a given discipline. Talking science means communicating in the language of science: academic English (Lemke, 1991). Learning and communicating in academic English poses challenges for many learners, including Eng+ students who may have less experience with the expected communication patterns in science. For example, in contexts like exams where students may be asked to articulate an explanation or argument for why a phenomenon occurs, Eng+ students may engage in additional cognitive operations (*e.g.*, translating questions) compared to English-as-a-first language (Eng1st) students to interpret English questions and then generate a written response in English (Abedi, 2002, 2015; Abedi et al., 2004; Abedi & Lord, 2001; Buxton et al., 2014; González-Howard & McNeill, 2016; E. N. Lee, 2018; E. N. Lee et al., 2020; E. N. Lee & Orgill, 2021; Solano-Flores & Trumbull, 2003).

The abovementioned linguistic challenges faced by Eng+ students can manifest into inequities in how Eng+ students are assessed and evaluated in science classrooms and educational research. Eng+ students' achievement in STEM fields has been found to be associated with language ability, regardless of students' abilities to grasp and apply scientific concepts (Curtis & Millar, 1988; Maerten-Rivera et al., 2010; Solano-Flores & Trumbull, 2003). Eng+ students' written responses can score lower on evaluations than those of Eng1st students due to evaluators' implicit bias (Huang, 2008; Lindsey & Crusan, 2011; Milnes & Cheng, 2008) and students' challenges communicating their knowledge (Lyon et al., 2012; Swanson et al., 2014; Wolf et al., 2008). From Kieffer et al. (2009): 'Because language plays an integral role in most, if not all, academic learning, any test of academic achievement is also, to some degree, a test of language ability' (Kieffer et al., 2009).

Neglecting the issues facing Eng+ students can result in increased attrition of students from a diversity of linguistic, cultural, and ethnic and backgrounds, and curbed development of key skills, such as scientific argumentation (Abedi, 2002; Afitska & Heaton, 2019; Allensworth, 2005; Buxton et al., 2014, 2015; Chin, 2010; Maerten-Rivera et al., 2010; McNeil, 2005; Pyburn

et al., 2013; Solano-Flores & Trumbull, 2003; Sotelo-Dynega et al., 2013; The National Academy of Sciences et al., 2018; Wolf et al., 2008; Yore et al., 2004). Such consequences may then limit the benefit that scientific and other professional communities might gain from individuals from Eng+ backgrounds (Amano et al., 2016, 2023; Cheng et al., 2019; Elnathan, 2021; Ramírez-Castañeda, 2020).

Without an understanding of how language might be associated with how students reason in chemistry, we are left wondering about whether opportunities for students to develop and demonstrate their reasoning skills are equitable. Much of the current work on language and learning has been conducted in K–12 mathematics, biology, or general science contexts, with few investigations in chemistry (Pyburn et al., 2013) and post-secondary contexts (LaCosse et al., 2020), despite chemistry's dependence on Anglophonic nomenclature, models, and symbols (Childs et al., 2015). Educational researchers have studied students' scientific reasoning and argumentation skills in various contexts and disciplines (Becker et al., 2016; Bodé et al., 2019; M. M. Cooper et al., 2016; Deng & Flynn, 2021; Moon et al., 2016, 2019; Osborne & Patterson, 2011; Russ et al., 2008; Watts et al., 2020; M. L. Weinrich & Talanquer, 2016); however, much of this work has relied on analyses of arguments constructed in a single language (usually English), without considering the role of language in these types of investigations. Therefore, in the research described in Chapter 5, we investigated how English language proficiency and history was associated with the reasoning in students' written chemistry arguments.

1.3.2 Linguistic equity in chemistry research

Much of the work on linguistic equity in science and chemistry education has focused on undergraduate students, with much less focused on graduate students and postdoctoral researchers (trainees). The limited focus on trainees is concerning given recent studies on the barriers faced by equity-deserving groups within graduate programs (Jaquette et al., 2016; Posselt & Grodsky, 2017). When discussing ways to improve EDI in chemistry, a focus on graduate students will be

imperative to enriching EDI within the broader chemistry community. Graduate students are the future educators, researchers, leaders, and professionals in our shared chemistry community, and promoting EDI within the graduate student population may bear positive downstream impacts on the state of EDI in our broader chemistry community.

The dominance of English is present not only in science classrooms, but also in research. A shared language like English facilitates communication and globalization of science, but it may also limit science as many students and researchers today learn English-as-an-additional language (Eng+) (Amano et al., 2016, 2023; Ramírez-Castañeda, 2020). For example, a study found that Eng+ biologists in Colombia faced significant barriers to publishing and presenting their research in English (Ramírez-Castañeda, 2020). Other work has described challenges with linguistic insecurity (feelings of anxiety) and accent discrimination amongst Eng+ students and scholars (Cheng et al., 2019; Chin, 2010; Lippi-Green, 1994; Rubin, 1992; Subtirelu, 2015). Within the scientific publication system, structural inequities exist that favour English publications and journals (Amano et al., 2016, 2021; Ammon, 2011; Angulo et al., 2021; Ramírez-Castañeda, 2020). In this thesis, I define linguistic equity as the removal of language barriers and biases that prevent equal access to a program (Government of Canada, 2020). The use of English in science is not inherently inequitable; however, the absence of resources and strategies to support equal participation in research can be considered inequitable.

With ongoing efforts to promote and support diversity in the chemical sciences, the effects of English as an international language on a chemistry community with rich linguistic diversity must be considered, as language dictates how we **learn**, **communicate**, and **do** science (Amano et al., 2023; Deng et al., 2022; Elsherief & Masson, 2020; E. N. Lee et al., 2020; O. Lee & Fradd, 1998). Deeper understanding of the role of language in science may also lead to a more holistic understanding of the experiences of racialized chemists, creating pathways towards equity, inclusion, and justice in science more broadly (Iftikar & Museus, 2018; Museus & Iftikar, 2013; Valdes, 1996).

Chapters 6 and 7 describe our research on the experiences of chemistry researchers from diverse language backgrounds, including trainees and faculty. Though there has been previous research on Eng+ learners in chemistry education (Flores & Smith, 2013; E. N. Lee & Orgill, 2021) this study is the first to focus on Eng+ chemistry trainees and faculty. Many chemistry researchers come from international contexts, playing a key role in sustaining and strengthening our future economies, environment, security, and health, but whose educational experiences are relatively understudied. Through a better understanding of Eng+ chemists' experiences, our community can be better equipped to support these chemists through their studies and careers, making chemistry more equitable and inclusive and enhancing scientific and economic progress (Amano et al., 2016; Angulo et al., 2021; Hofstra et al., 2020).

1.4 Thesis overview

This thesis is manuscript-based. Each chapter in this thesis focuses on a published, peer-reviewed manuscript describing research that I led or contributed to during my doctoral studies (with the exception of Chapter 7, which is a manuscript in preparation). Each chapter starts with brief introductions to connect between chapters and detail the key issues/goals informing the research in each chapter. Also included are brief personal reflections on my own intellectual and professional development at the time of each project in hopes of demonstrating how my research expertise and interests evolved over the course of my degree. I was and am still learning, and I want to make evident to readers how expertise, self-efficacy, and interests can evolve over the course of one's studies, career, and life.

Chapters 2–8 describe projects and manuscripts I contributed to or led as a member of the Flynn Research Group (i.e., projects in collaboration with my supervisor Prof. Alison Flynn and other Flynn group members). Chapters 9–11 describe projects and manuscripts I contributed to or led outside of the Flynn group—that is, projects I contributed to or led as an independent researcher and scholar with colleagues from outside the Flynn group ("external projects"). Supporting Information for each re-printed manuscript is included in the Appendices chapter (Chapter 13).

I include the Chapter 9–11 projects in my thesis for three reasons. First, many of the ideas explored in these external projects overlap with and partially informed (or were informed by) the ideas and projects I contributed to in the Flynn group. For example, "Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities" (Chapter 9) contributed in part to my interest in pursuing a doctoral research project on "The experiences of chemistry trainees learning English-as-an-additional language" (Chapter 6), from which the findings contributed to my interest in and leadership on the editorial "Prioritizing Mentorship as Scientific Leaders" (Chapter 11).

Second, these projects gave me amazing opportunities to grow as a researcher and contribute to the scholarly community in ways that complemented or supplemented my work with the Flynn group. These external projects allowed me to work with researchers, administrators, educators, and policymakers outside of our group, department, research field, and beyond, collaborating with them on different projects at various scales (local, national, international). These projects have deeply shaped my learning, expertise, and interests, and I view them as essential evidence of my development and contributions as a researcher and academic.

Lastly, many of these projects emerged from participation in committees and organizations that might be viewed traditionally as service work—work that is historically "invisible" within an academic system primarily interested in research output and productivity (Social Sciences Feminist Network Research Interest Group, 2017; Truong, 2021). I choose to include these projects in the hopes that their inclusion contributes to ongoing efforts to normalize and re-orient how we choose to value service and other "non-research" work in academia and science.

1.5 References for Chapter 1

Abedi, J. (2002). Assessment and accommodations of English language: Issues, concerns, and recommendations. *Journal of School Improvement*, *3*(1), 83–89.

Abedi, J. (2015). Language Issues in Item Development. *Handbook of Test Development*, *Civ*,

355–373. https://doi.org/10.4324/9780203874776.ch17

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for english language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*(1), 1–28. https://doi.org/10.3102/00346543074001001

Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*, *14*(3), 219–234. https://doi.org/10.1207/S15324818AME1403

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining Effectiveness and Validity of Accommodations for English Language Learners in Mathematics: An Evidence-Based Computer Accommodation Decision System. *Educational Measurement: Issues and Practice*, *39*(4), 41–52. https://doi.org/10.1111/emip.12328

Abrams, E., & Southerland, S. (2010). The how's and why's of biological change : How learners neglect physical mechanisms in their search for meaning. *International Journal of Science Education*, *23*(12), 1271–1281. https://doi.org/10.1080/09500690110038558

Afitska, O., & Heaton, T. J. (2019). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, *103*(6), 1396–1422. https://doi.org/10.1002/sce.21545

Allensworth, E. (2005). Graduation and Dropout Trends in Chicago: A Look at Cohorts of Students from 1991 through 2004. In *Consortium on Chicago School Research* (Issue January). papers://e97baf73-a10d-438d-985b-bee59744e11f/Paper/p939

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages Are Still a Major Barrier to Global Science. *PLoS Biology*, *14*(12), e2000933. https://doi.org/10.1371/journal.pbio.2000933

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., David González-Trujillo, J., Montaño-3 Centellas, F., Paudel, K., White, R., & Veríssimo, D. (2023). The cost of being a non-native English speaker in science. *PLoS ONE*, *21*(7), 1–27.

Amano, T., Rios Rojas, C., Boum II, Y., Calvo, M., & Misra, B. B. (2021). Ten tips for overcoming language barriers in science. *Nature Human Behaviour*. https://www.nature.com/articles/s41562-021-01137-1

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Ammon, U. (Ed.). (2011). *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110869484

Angulo, E., Diagne, C., Ballesteros-Mejia, L., Adamjy, T., Ahmed, D. A., Akulov, E., Banerjee, A. K., Capinha, C., Dia, C. A. K. M., Dobigny, G., Duboscq-Carra, V. G., Golivets, M., Haubrock, P. J., Heringer, G., Kirichenko, N., Kourantidou, M., Liu, C., Nuñez, M. A., Renault, D., … Courchamp, F. (2021). Non-English languages enrich scientific knowledge: The example of economic costs of biological invasions. *Science of the Total Environment*, *775*, 144441. https://doi.org/10.1016/j.scitotenv.2020.144441

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, *90*(5), 536–545. https://doi.org/10.1021/ed3002013

Atkinson, M. B., Popova, M., Croisant, M., Reed, D. J., & Bretz, S. L. (2020). Development of the Reaction Coordinate Diagram Inventory: Measuring Student Thinking and Confidence. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.9b01186

Barrett, L. F., & Bliss-Moreau, E. (2009). She's Emotional. He's Having a Bad Day: Attributional Explanations for Emotion Stereotypes. *Emotion*, *9*(5), 649–658.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448. https://doi.org/10.1007/s11218-012-9185-3

Becker, N., Noyes, K., & Cooper, M. M. (2016). Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *Journal of Chemical Education*, *93*(10), 1713–1724. https://doi.org/10.1021/acs.jchemed.6b00298

Berland, L. K., & Reiser, B. J. (2009). Making Sense of Argumentation and Explanation. *Science Education*, *93*, 26–55. https://doi.org/10.1002/sce.20286

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Bodner, G. M., & Orgill, M. (Eds.). (2007). *Theoretical Frameworks for Research in Chemistry/Science Education*. Pearson.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*(February 2002), 605–634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Breda, T., Grenet, J., Monnet, M., & Effenterre, C. van. (2018). Can female role models reduce the gender gap in science ? Evidence from classroom interventions in French high schools. *Hal*, *October*, 18–19. https://halshs.archives-ouvertes.fr/halshs-01713068

Bunce, D. M. (2008). Constructing good and researchable questions. In *Nuts and Bolts of Chemical Education Research* (pp. 35–46). https://doi.org/10.1021/bk-2008-0976.ch004

Buxton, C. A., Allexsaht-Snider, M., Aghasaleh, R., Kayumova, S., Kim, S.-H., Choi, Y.-J., & Cohen, A. (2014). Potential Benefits of Bilingual Constructed Response Science Assessments for Understanding Bilingual Learners' Emergent Use of Language of Scientific Investigation Practices. *Double Helix*, *2*(September 2016), 1–21.

Buxton, C. A., Salinas, A., Mahotiere, M., Lee, O., & Secada, W. (2015). Fourth-grade emergent bilingual learners' scientific reasoning complexity, controlled experiment practices, and content knowledge when discussing school, home, and play contexts. *Teachers College*

Record, *117*(2).

Campbell, C. E., & Nehm, R. H. (2013). A critical analysis of assessment quality in genomics and bioinformatics education research. *CBE Life Sciences Education*, *12*(3), 530–541. https://doi.org/10.1187/cbe.12-06-0073

Caspari, I, Kranz, D., & Graulich, N. (2018). Resolving the complexity of organic chemistry students ' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.*, *19*, 1117–1141. https://doi.org/10.1039/C8RP00131F

Caspari, Ira, Weinrich, M. L., Sevian, H., & Graulich, N. (2018). This mechanistic step is "productive": organic chemistry students' backward-oriented reasoning. *Chemistry Education Research and Practice*, *19*(1), 42–59. https://doi.org/10.1039/c7rp00124j

Cheng, Y., Dharwadkar, S., Sheridan, V., Rojas, C. R., Bosch Grau, M., & Gordin, M. (2019). Science's language barrier. *Nature*, *570*, 265.

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, *6*(FEB), 1–8. https://doi.org/10.3389/fpsyg.2015.00049

Childs, P. E., Markic, S., & Ryan, M. C. (2015). The Role of Language in Teaching and Learning Chemistry. In J. Garcia-Martinez & E. Serrano-Torregrossa (Eds.), *Chemistry Education: Best Practices, Opportunities and Trends* (pp. 421–445). Wiley-VCH Verlag GmbH & Co. KGaA.

Chin, W. Y. (2010). Linguistic profiling in education: how accent bias denis equal educational opportunites to students of color. In *The Scholar* (Vol. 12, Issue 355, pp. 356–384).

Cooper, A. (2016). *Systematic planning for knowledge mobilization: moving beyond ad hoc efforts towards intentional impact*. http://www.ripplenetwork.ca/wp-content/uploads/2019/03/Cooper 2016 Toolkit Approaches to KMb Planning.pdf

Cooper, M. M., & Klymkowsky, M. (2013). Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical*

Education, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating Students' Reasoning about Acid-Base Reactions. *Journal of Chemical Education*, *93*(10), 1703–1712. https://doi.org/10.1021/acs.jchemed.6b00417

Cooper, M. M., & Stowe, R. L. (2018). Chemistry Education Research - From Personal Empiricism to Evidence, Theory, and Informed Practice. *Chemical Reviews*, *118*(12), 6053–6087. https://doi.org/10.1021/acs.chemrev.8b00020

Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2019). Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *Journal of Chemical Education*, *96*(2), 213–226. https://doi.org/10.1021/acs.jchemed.8b00784

Crandell, O. M., Lockhart, M. A., & Cooper, M. M. (2020). Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry. *Journal of Chemical Education*, acs.jchemed.9b00815. https://doi.org/10.1021/acs.jchemed.9b00815

Creswell, J.W., & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. *Theory Into Practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903

Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 443–507). American Council on Education.

Crossman, E., Lu, Y., & Hou, F. (2022). International students as a source of labour supply: Engagement in the labour market after graduation. *Economic and Social Reports*, *2*(3), 1–5.

Curran, M., & Randall, A. K. (2020). *Positionality Statements*. International Association for Relationship Research. https://doi.org/10.1177/1745691620927709

Curtis, S., & Millar, R. (1988). Language and Conceptual Understanding in Science: A comparison of English and Asian language speaking children. *Research in Science &*

Technological Education, *6*(1), 61–77. https://doi.org/10.1080/0263514880060106

Darden, L. (2002). Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. *Philosophy of Science*, *69*(S3), 354–365.

Darwin Holmes, A. G. (2020). Researcher Positionality - A Consideration of Its Influence and Place in Qualitative Research - A New Researcher Guide. *Shanlax International Journal of Education*, *8*(4), 1–10. https://doi.org/10.34293/education.v8i4.3232

DeCocq, V., & Bhattacharyya, G. (2019). TMI (Too much information)! Effects of given information on organic chemistry students' approaches to solving mechanism tasks. *Chemistry Education Research and Practice*, *20*(1), 213–228. https://doi.org/10.1039/C8RP00214B

Deng, J. M., & Flynn, A. B. (2021). Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items. *Chemistry Education Research and Practice*, *22*, 749–771. https://doi.org/10.1039/D0RP00320D

Deng, J. M., Rahmani, M., & Flynn, A. B. (2022). The role of language in students' justifications of scientific phenomena. *International Journal of Science Education*, *44*(13), 2131–2151. https://doi.org/10.1080/09500693.2022.2114299

Deng, J. M., Streja, N., & Flynn, A. B. (2021). Response Process Validity Evidence in Chemistry Education Research. *Journal of Chemical Education*, *98*(12), 3656–3666. https://doi.org/10.1021/acs.jchemed.1c00749

Denofrio, L. A., Russell, B., Lopatto, D., & Lu, Y. (2007). Mentoring: Linking student interests to science curricula. *Science*, *318*(5858), 1872–1873. https://doi.org/10.1126/science.1150788

Elnathan, R. (2021). *English is the language of science — but precision is tough as a non-native speaker*. Nature. https://www.nature.com/articles/d41586-021-00899-y

Elsherief, H., & Masson, M. (2020). *Every Teacher is a Language Teacher* (Vol. 1). University of Ottawa Second Language Education Cohort (cL2c).

https://doi.org/http://dx.doi.org/10.20381/an3d-6k16

Etkin, N., Armour, M.-A., Franz-Odendaal, T., Rayner-Canham, G., Rayner-Canham, M., Al-Abadleh, H. A., & Jones, K. (2016). *Equity and Diversity in Chemistry* (N. Etkin (Ed.)). MacMillan Learning Curriculum Solutions.

European Union. (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. *Official Journal of the European Union*, L 394/19-L 394/18.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Flores, A., & Smith, K. C. (2013). Spanish-speaking english language learners' experiences in high school chemistry education. *Journal of Chemical Education*, *90*(2), 152–158. https://doi.org/10.1021/ed300413j

Flynn, A. B., & Featherstone, R. B. (2017). Language of mechanisms: exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions. *Chem. Educ. Res. Pract.*, *18*(1), 64–77. https://doi.org/10.1039/C6RP00126B

Flynn, A., Campbell-Brown, E., O'Connor, E., Walsh, E., O'Hagan, F., Richard, G., & Roy, K. (2020). *Growth & Goals: a course-integrated module to better equip students with learning skills*. https://ecampusontario.pressbooks.pub/growthandgoalscourse/

Flynn, D. T. (2016). STEM Field Persistence: The Impact of Engagement on Postsecondary STEM Persistence for Underrepresented Minority Students. *Journal of Educational Issues*, *2*(1), 185. https://doi.org/10.5296/jei.v2i1.9245

Fuentes, E., & Liou, D. (2008). Language, Culture, and Access. In P. A. Nogueira & J. Y. Wing (Eds.), *Unfinished Business: Closing the Racial Achievement Gap in Our Schools* (pp. 52–68). Jossey-Bass.

Funk, C., & Parker, K. (2018). Women and Men in STEM Often at Odds Over Workplace

Equity. *Pew Research Center*, *January*, 1–19. http://www.pewsocialtrends.org/2018/01/09/women-and-men-in-stem-often-at-odds-over-workplace-equity/#fn-24050-6

Galloway, K. R., Leung, M. W., & Flynn, A. B. (2019). Patterns of reactions: A card sort task to investigate students' organization of organic chemistry reactions. *Chemistry Education Research and Practice*, *20*(1), 30–52. https://doi.org/10.1039/c8rp00120k

Galloway, K. R., Stoyanovich, C., & Flynn, A. B. (2017). Students' interpretations of mechanistic language in organic chemistry before learning reactions. *Chem. Educ. Res. Pract.*, *18*(2), 353–374. https://doi.org/10.1039/C6RP00231E

Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *Internet and Higher Education*, *9*(1), 1–8. https://doi.org/10.1016/j.iheduc.2005.11.001

González-Howard, M., & McNeill, K. L. (2016). Learning in a community of practice: Factors impacting english-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, *53*(4), 527–553. https://doi.org/10.1002/tea.21310

González-Pérez, S., Mateos de Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is It a Female Role-Model Thing? *Frontiers in Psychology*, *11*(September). https://doi.org/10.3389/fpsyg.2020.02204

Government of Canada. (2020). *Best Practices in Equity, Diversity and Inclusion in Research*. https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/edi-eng.aspx

Government of Canada Tri-Agencies. (2022). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2*. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: time for a map? *The Journal of Continuing Education in the Health Professions*, *26*(1), 13–24. https://doi.org/10.1002/chp.47

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor. Quant. Methods. Psychol.*, *8*(1), 23–34.

Hesse-Biber, S. (2010). Qualitative approaches to mixed methods practice. *Qualitative Inquiry*, *16*(6), 455–468. https://doi.org/10.1177/1077800410364611

Higher Education Strategy Associates. (2020). *The state of post-secondary education in Canada*.

Hofstra, B., Kulkarni, V. V., Galvez, S. M. N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Holme, T. A., & Hutchison, J. E. (2018). A Central Learning Outcome for the Central Science. *Journal of Chemical Education*, *95*(4), 499–501. https://doi.org/10.1021/acs.jchemed.8b00174

Howe, M. E., Kim, M. M., & Pazicni, S. (2022). Graduate Student Women's Perceptions of Faculty Careers: The Critical Role of Departmental Values and Support in Career Choice. *Journal of the American Chemical Society Au*. https://doi.org/10.1021/jacsau.2c00175

Howell, E. L., & Brossard, D. (2021). (Mis)informed about what? What it means to be a science-literate citizen in a digital world. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(15), 1–8. https://doi.org/10.1073/pnas.1912436117

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments?-A generalizability theory approach. *Assessing Writing*, *13*(3), 201–218. https://doi.org/10.1016/j.asw.2008.10.002

Iftikar, J. S., & Museus, S. D. (2018). On the utility of Asian critical (AsianCrit) theory in the field of education. *International Journal of Qualitative Studies in Education*, *31*(10), 935–949. https://doi.org/10.1080/09518398.2018.1522008

Institute of International Education. (2020). *Project Atlas: A Quick Look at Global Mobility Trends*. https://iie.widen.net/s/g2bqxwkwqv/project-atlas-infographics-2020

Isgro, K., & Castaneda, M. (2015). Mothers in U.S. academia: Insights from lived experiences. *Women's Studies International Forum*, *53*, 174–181.

Jaquette, O., Curs, B. R., & Posselt, J. R. (2016). Tuition Rich, Mission Poor: Nonresident Enrollment Growth and the Socioeconomic and Racial Composition of Public Research Universities. *The Journal of Higher Education*, *87*(5), 635–673. https://doi.org/10.1080/00221546.2016.11777417

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12001

Kelemen, D., & Rosset, E. (2009). The Human Function Compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143. https://doi.org/10.1016/j.cognition.2009.01.001

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. In *Review of Educational Research* (Vol. 79, Issue 3). https://doi.org/10.3102/0034654309332490

Kolodner, J. L. (1992). An Introduction to Case-Based Reasoning. *Artificial Intelligence Review*, *6*, 3–34. https://doi.org/10.1136/bmj.4.5576.398

Kraft, A., Strickland, A. M., & Bhattacharyya, G. (2010). Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chemistry Education Research and Practice*, *11*(4), 281–292. https://doi.org/10.1039/C0RP90003F

Kredell, M. (2017). Equity Runs Through It. *USC Rossier*, 20–23.

Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific Reasoning in Science Education: From Global Measures to Fine-Grained Descriptions of Students' Competencies. *Education Sciences*, *12*(97), 1–8.

Krippendorff, K. (1970). Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educational and Psychological Measurement*, *30*(1), 61–70.

LaCosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-

belonging intervention improves STEM outcomes for students who speak English as a second language. *Science Advances*, *6*(40), 1–11. https://doi.org/10.1126/sciadv.abb6543

Lapierre, K. (2019). *Exploring the influence of a redesigned organic chemistry curriculum on students' interpretations of reactions and self-efficacy beliefs in organic chemistry*. University of Ottawa.

Lavis, J. N., Robertson, D., Woodside, J. M., McLeod, C. B., & Abelson, J. (2003). How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Quarterly*, *81*(2), 221–248. https://doi.org/10.1111/1468-0009.t01-1-00052

Lee, E. N. (2018). *Breaking the language barrier: Equitable assessment in general chemistry (doctoral dissertation)* [University of Nevada, Las Vegas]. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2018-58618-069&lang=de&site=ehost-live

Lee, E. N., & Orgill, M. (2021). Toward Equitable Assessment of English Language Learners in General Chemistry: Identifying Supportive Features in Assessment Items. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.1c00370

Lee, E. N., Orgill, M., & Kardash, C. (2020). Supporting English Language Learners in College Science Classroom: Insights from Chemistry Students. *Multicultural Education*, *27*(Spring/Summer), 25–32.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-Language backgrounds. *Educational Researcher*, *27*(4), 12–21. https://doi.org/10.3102/0013189X027004012

Lee, O., Quinn, H., & Valdés, G. (2013). Science and Language for English Language Learners in Relation to Next Generation Science Standards and with Implications for Common Core State Standards for English Language Arts and Mathematics. *Educational Researcher*, *42*(4), 223–233. https://doi.org/10.3102/0013189X13480524

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the Starting Gate: Social Background*

Differences in Achievement as Children Begin School.

Lemke, J. L. (1991). Talking science: language, learning, and values. In *Choice Reviews Online* (Vol. 28, Issue 09). https://doi.org/10.5860/choice.28-5211

Lindsey, P., & Crusan, D. (2011). How Faculty Attitudes and Expectations Toward Student Nationality Affect Writing Assessment. *Across the Disciplines*, *8*(4), 1–19. https://doi.org/10.37514/atd-j.2011.8.4.23

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, *23*(2), 163–198. https://doi.org/10.1017/S0047404500017826

Luisi, P. L. (2002). Emergence in Chemistry: Chemistry as the Embodiment of Emergence. *Foundations in Chemistry*, *4*(3), 183–200.

Lyon, E. G., Bunch, G. C., & Shaw, J. M. (2012). Navigating the language demands of an inquiry-based science performance assessment: Classroom challenges and opportunities for English learners. *Science Education*, *96*(4), 631–651. https://doi.org/10.1002/sce.21008

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education*, *94*(6), 937–962. https://doi.org/10.1002/sce.20408

Maeyer, J., & Talanquer, V. (2013). Making Predictions About Chemical Reactivity: Assumptions and Heuristics. *Journal of Research in Science Teaching*, *50*(6), 748–767. https://doi.org/10.1002/tea.21092

Maxwell, J. A. (2012). *Qualitative Research Design: An Interactive Approach*. SAGE Publications Inc.

McClary, L., & Talanquer, V. (2011). Heuristic Reasoning in Chemistry: Making decisions about acid strength. *International Journal of Science Education*, *33*(10), 1433–1454.

McNeil, L. (2005). Faking equity: High-stakes testing and the education of Latino youth. In A.

Valanzeula (Ed.), *Leaving children behind: How "Texas style" accountability fails Latino youth* (pp. 57–112). University of New York Press.

McNeill, K. L., & Berland, L. K. (2010). A Learning Progression for Scientific Argumentation: Understanding Student Work and Designing Supportive. *Science Education*, 765–793. https://doi.org/10.1002/sce.20402

McNeill, K. L., & Krajcik, J. (2008a). Inquiry and Scientific Explanations: Helping Students Use Evidence and Reasoning. In *Science as Inquiry in the Secondary Setting* (pp. 121–134).

McNeill, K. L., & Krajcik, J. (2008b). Scientific Explanations: Characterizing and Evaluating the Effects of Teachers' Instructional Practices on Student Learning. *Journal of Research in Science Teaching*, *45*(1), 53–78. https://doi.org/10.1002/tea

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 741–749.

Milnes, T., & Cheng, L. (2008). Teachers' Assessment of ESL Students in Mainstream Classes: Challenges, Strategies, and Decision-Making. *TESL Canada Journal*, *26*(1), 49. https://doi.org/10.18806/tesl.v26i1.129

Moon, A., Moeller, R., Gere, A. R., & Shultz, G. V. (2019). Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification. *Chemistry Education Research and Practice*, *20*(3), 484–494. https://doi.org/10.1039/c9rp00005d

Moon, A., Stanford, C., Cole, R., & Towns, M. (2016). The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry. *Chemistry Education*

Research and Practice, *17*(2), 353–364. https://doi.org/10.1039/c5rp00207a

Moss-Racusin, C. A., & Molenda, A. K. (2015). Can Evidence Impact Attitudes? Public Reactions to Evidence of Gender Bias in STEM Fields. *Psychology of Women Quarterly*, *39*(2), 194–209.

Museus, S. D., & Iftikar, J. (2013). An Asian Critical Theory (AsianCrit) Framework. *Asian American Students in Higher Education*, 18–29.

National Academies of Science Engineering and Medicine. (2018). *Graduate STEM Education for the 21st Century* (A. Leshner & L. Scherer (Eds.)). National Academies Press. https://doi.org/10.17226/25038

National Research Council. (2012a). *A Framework for K-12 Science Education*. National Academies Press. https://www.nap.edu/read/13165/chapter/1

National Research Council. (2012b). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. In *The National Academies Press*. https://doi.org/10.17226/13362

National Research Council. (2013). *Next Generation Science Standards*. https://www.nextgenscience.org/

National Science Foundation. (2016). *Doctoral Scientists and Engineers in Academia*. National Science Board: Science and Engineering Indicators. https://www.nsf.gov/statistics/2016/nsb20161/#/report/chapter-5/doctoral-scientists-and-engineers-in-academia/federal-support-of-doctoral-researchers-in-academia

National Science Foundation. (2017). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017 Digest Special Report NSF 17-310.* https://www.nsf.gov/statistics/2017/nsf17310/static/downloads/nsf17310-digest.pdf

National Science Foundation. (2020). *2019 Doctorate Recipients from U.S. Universities: Vol. NSF 21-308* (Issue December 2020). http://www.nsf.gov/statistics/2016/nsf16300/digest/nsf16300.pdf

Nersessian, N. J. (1999). Model-Based Reasoning in Conceptual Change. In *Model-Based Reasoning in Scientific Discovery* (pp. 5–22). https://doi.org/10.1007/978-1-4615-4813-3_1

O'Brian, L. T., Blodorn, A., Adams, G., & Garcia, D. M. (2015). Ethnic Variation in Gender-STEM Stereotypes and STEM Participant: An Intersectional Approach. *Cultural Diversity and Ethnic Minority Psychology*, *21*(2), 169–180.

O'Connor, E., Roy, K., Walsh, E., Huang, D., Ke, D. Y. J., Campbell Brown, E., Moreau, K. A., & Flynn, A. B. (2021). *The Evaluation of an Integrated Growth & Goals Module to Better Equip Students with Learning Skills in Postsecondary Courses: Systematic, Scalable, and Explicit.* https://doi.org/doi:10.26434/chemrxiv.13543130.v2

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, *49*(6), 744–777. https://doi.org/10.1002/tea.21028

Organisation for Economic Cooperation and Development. (2006). *Assessing scientific, reading and mathematical literacy: a framework for PISA 2006*.

Osborne, J. F., & Patterson, A. (2011). Scientific Argument and Explanation: A Necessary Distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438

Osborne, J. F., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, *50*(3), 315–347. https://doi.org/10.1002/tea.21073

Posselt, J. R. (2020). *Equity in Science: Representation, Culture, and the Dynamics of Change in Graduate Education*. Stanford University Press. http://www.sup.org/books/title/?id=31068

Posselt, J. R., & Grodsky, E. (2017). Graduate Education and Social Stratification. *Annual Review of Sociology*, *43*, 353–378. https://doi.org/10.1146/annurev-soc-081715-074324.Graduate

Pyburn, D. T., Pazicni, S., Benassi, V. A., & Tappin, E. E. (2013). Assessing the relation between

language comprehension and performance in general chemistry. *Chemistry Education Research and Practice*, *14*(4), 524–541. https://doi.org/10.1039/c3rp00014a

Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS ONE*, *15*(9), 1–15. https://doi.org/10.1371/journal.pone.0238372

Reuben, E., & Sapienza, P. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4403–4408.

Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers* (2nd ed.). Blackwell Publishing Professional.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of non-native English-speaking teaching assistants. *Research in Higher EducationEducation*, *33*, 511–531.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed From Philosophy of Science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

Sadler, T. D., Barab, S. A., & Scott, B. (2007). What Do Students Gain by Engaging in Socioscientific Inquiry? *Research in Science Education*, *37*(4), 371–391. https://doi.org/10.1007/s11165-006-9030-9

Sanford, M. (2020). Equity and Inclusion in the Chemical Sciences Requires Actions not Just Words. *Journal of the American Chemical Society*, 4–5. https://doi.org/10.1021/jacs.0c06482

Sarsons, H. (2015). *Gender Differences in Recognition for Group Work.*

Schwartz, D. L. (1995). The Emergence of Abstract Representations in Dyad Problem Solving. *Journal of the Learning Sciences*, *4*(3), 321–354.

https://doi.org/10.1207/s15327809jls0403_3

Sevian, H., Bernholt, S., Szteinberg, G. A., & Auguste, S. (2015). Use of representation mapping to capture abstraction in problem solving in different courses in chemistry. *Chemistry Education Research and Practice*, *16*(3), 429–446. https://doi.org/10.1039/c5rp00030k

Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: a learning progression on chemical thinking. *Chem. Educ. Res. Pract.*, *15*(1), 10–23. https://doi.org/10.1039/c3rp00111c

Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify misinformation in everyday life? *Science Education*, *September 2018*, 1–22. https://doi.org/10.1002/sce.21581

Sirum, K., & Humburg, J. (2011). The experimental design ability test (EDAT). *Bioscene*, *37*(1), 9–16.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi-org.proxy.bib.uottawa.ca/10.1037/0033-2909.119.1.3

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate stem (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The genetics concept assessment: A new concept inventory for gauging student understanding of genetics. *CBE Life Sciences Education*, *7*(4), 422–430. https://doi.org/10.1187/cbe.08-08-0045

Social Sciences and Humanities Research Council. (2018). *Truth Under Fire in a Post-Fact World*. https://horizons.gc.ca/en/2018/10/19/the-next-generation-of-emerging-global-challenges/#truth-under-fire

Social Sciences and Humanities Research Council. (2019). *Guidelines for Effective Knowledge Mobilization*. https://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/knowledge_mobilisation-mobilisation_des_connaissances-eng.aspx

Social Sciences Feminist Network Research Interest Group. (2017). The Burden of Invisible Work in Academia: Social Inequalities and Time Use in Five University Departments. *Humboldt Journal of Social Relations*, *39*(39), 228–245. https://ci.nii.ac.jp/naid/40021243259/

Solano-Flores, G., & Trumbull, E. (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners. *Educational Researcher*, *32*(2), 3–13. https://doi.org/10.3102/0013189X032002003

Sotelo-Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. (2013). English language proficiency and test performance: an evaluation of bilingual students with the Woodcock-Johnson III Tests of Cogntitive Abilities. *Psychology in the Schools*, *50*(8), 781–797. https://doi.org/10.1002/pits

Southard, K. M., Espindola, M. R., Zaepfel, S. D., & Molly, S. (2017). Generative mechanistic explanation building in undergraduate molecular and cellular biology. *International Journal of Science Education*, *39*(13), 1795–1829. https://doi.org/10.1080/09500693.2017.1353713

Sternberg, R. J. (1986). Toward a unified theory of human reasoning. *Intelligence*, *10*(4), 281–314. https://doi.org/10.1016/0160-2896(86)90001-2

Subtirelu, N. C. (2015). She does have an accent but: Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society*, *44*(1), 35–62. https://doi.org/10.1017/S0047404514000736

Swanson, L. H., Bianchini, J. A., & Lee, J. S. (2014). Engaging in argument and communicating information: A case study of english language learners and their science teacher in an urban high school. *Journal of Research in Science Teaching*, *51*(1), 31–64. https://doi.org/10.1002/tea.21124

Taber, K. S. (2013). Three levels of chemistry educational research. *Chemistry Education Research and Practice*, *14*(2), 151–155. https://doi.org/10.1039/c3rp90003g

Talanquer, V. (2007). Explanations and Teleology in Chemistry Education. *International Journal*

of Science Education, *29*(7), 853–870. https://doi.org/10.1080/09500690601087632

Talanquer, V. (2017). Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, *94*(12), 1805–1810. https://doi.org/10.1021/acs.jchemed.7b00427

Talanquer, V. (2018a). Assessing for Chemical Thinking. In *Research and Practice in Chemistry Education* (pp. 123–133). Springer Nature Singapore Pte Ltd. 2019. https://doi.org/10.1007/978-981-13-6998-8_8

Talanquer, V. (2018b). Progressions in reasoning about structure – property relationships. *Chemistry Education Research and Practice*, *19*(4), 998–1009. https://doi.org/10.1039/c7rp00187h

Talanquer, V. (2018c). Progressions in reasoning about structure–property relationships. *Chemistry Education Research and Practice*, *19*(4), 998–1009. https://doi.org/10.1039/C7RP00187H

Talanquer, V., & Pollard, J. (2010). Let's teach how we think instead of what we know. *Chemistry Education Research and Practice*, *11*(2), 74–83. https://doi.org/10.1039/c005349j

The National Academy of Sciences, National Academies of Science Engineering and Medicine, & The National Academy of Sciences. (2018). English Learners in STEM Subjects: Transforming Classrooms, Schools, and Lives. In D. Francis & A. Stephens (Eds.), *English Learners in STEM Subjects*. The National Academies Press. https://doi.org/10.17226/25182

The Royal Society of Chemistry. (2018). *Diversity landscape of the chemical sciences*. https://www.rsc.org/globalassets/02-about-us/our-strategy/inclusion-diversity/cm-044-17_a4-diversity-landscape-of-the-chemical-sciences-report_web-2.pdf

Tilghman, B. S., Alberts, B., Colón-ramos, D., Dzirasa, K., Kimble, J., & Varmus, H. (2021). Concrete steps to diversify the scientific workforce. *Science Magazine*, 133–136.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge University Press.

https://doi.org/10.1080/00048405985200191

Trommler, F., Gresch, H., Hammann, M., Trommler, F., Gresch, H., & Hammann, M. (2018). Students' reasons for preferring teleological explanations. *International Journal of Science Education*, *40*(2), 159–187. https://doi.org/10.1080/09500693.2017.1404658

Truong, K. A. (2021). *Making the Invisible Visible*. Inside Higher Ed. https://www.insidehighered.com/advice/2021/05/28/why-and-how-colleges-should-acknowledge-invisible-labor-faculty-color-opinion

Uhl Chamot, A. (1995). Implementing the Cognitive Academic Language Learning Approach: CALLA in Arlington, Virginia. *Bilingual Research Journal*, *19*(3 & 4), 379–394.

United Nations. (2015). *Transforming our World: the 2030 Agenda for Sustainable Development*. https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Valdes, F. (1996). Latina/o Ethnicities, Critical Race Theory, and Post-Identity Politics in Postmodern Legal Culture: From Practices to Possibilites Foreword. *La Raza L.J.*, *9*(1994), 1.

van Mil, M. H. W., Jan, D., Arend, B., & Waarlo, J. (2013). Modelling Molecular Mechanisms : A Framework of Scientific Reasoning to Construct Molecular-Level Explanations for Cellular Behaviour. *Science Education*, *22*(1), 93–118. https://doi.org/10.1007/s11191-011-9379-7

Ward, S. L., Byrnes, J. P., & Overton, W. F. (1990). Organization of Knowledge and Conditional Reasoning. *Journal of Educational Psychology*, *82*(4), 832–837.

Watts, F. M., Schmidt-McCormack, J. A., Wilhelm, C. A., Karlin, A., Sattar, A., Thompson, B. C., Gere, R., & Shultz, G. V. (2020). What students write about when students write about mechanisms: analysis of features present in students' written descriptions of an organic reaction mechanism. *Chemistry Education Research and Practice*, *21*, 1148–1172. https://doi.org/10.1039/c9rp00185a

Weinrich, M. L., & Talanquer, V. (2016). Mapping students' modes of reasoning when thinking

about chemical reactions used to make a desired product. *Chemistry Education Research and Practice*, *17*(2), 394–406. https://doi.org/10.1039/c5rp00208g

Weinrich, Melissa L., & Sevian, H. (2017). Capturing students' abstraction while solving organic reaction mechanism problems across a semester. *Chemistry Education Research and Practice*, *18*(1), 169–190. https://doi.org/10.1039/c6rp00120c

Whitcomb, K. M., & Singh, C. (2020). Not all disadvantages are equal: Racial/ethnic minority students have largest disadvantage of all demographic groups in both STEM and non-STEM GPA. *ArXiv*, *15260*.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Science Education*, *92*(5), 941–967. https://doi.org/10.1002/sce.20259

Winfield, L. L., Wilson-Kennedy, Z. S., Payton-Stewart, F., Nielson, J., Kimble-Hill, A. C., & Arriaga, E. A. (2020). Journal of Chemical Education Call for Papers: Special Issue on Diversity, Equity, Inclusion, and Respect in Chemistry Education Research and Practice. *Journal of Chemical Education*, *97*(11), 3915–3918. https://doi.org/10.1021/acs.jchemed.0c01300

Wisdom, J., & Creswell, J. W. (2013). Mixed Methods: Integrating Quantitative and Qualitative Data Collection and Analysis While Studying Patient-Centered Medical Home Models. *Agency for Healthcare Reseach and Quality*, *13-0028-EF*, 1–5. https://doi.org/No. 13-0028-EF.

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses. In *CRESST Report 732*.

Woolston, C. (2020). White men still dominate in UK academic science. *Nature*, *579*, 622. https://doi.org/10.2165/00128413-199208320-00046

Wren, D., & Barbera, J. (2013). *Gathering Evidence for Validity during the Design,*

Development, and Qualitative Evaluation of Thermochemistry Concept Inventory Items. https://doi.org/10.1021/ed400384g

Yoder, J. B., & Mattheis, A. (2016). Queer in STEM: Workplace Experiences Reported in a National Survey of LGBTQA Individuals in Science, Technology, Engineering, and Mathematics Careers. *Journal of Homosexuality*, *63*(1), 1–27.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly*, *39*(3), 347–352. https://doi.org/10.1598/RRQ.39.3.8

Youssef, A. (2023). *The language of organic chemistry: is fluency the key to success?* University of Ottawa.

Chapter 2: Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms

2.1 Introduction to the research article presented in this chapter

Constructing arguments using scientific evidence and reasoning is a key skill in a world facing complex global challenges (European Union, 2006; National Research Council, 2012a; Organisation for Economic Cooperation and Development, 2006; Social Sciences and Humanities Research Council, 2018). To equip learners with these core skills, efforts in chemistry education and beyond have incorporated opportunities for learners to practice and demonstrate their argumentation skills, such as in formative and summative assessments (*e.g.*, prompting students to make a claim and justify that claim in writing). With these opportunities, research is needed to understand how learners are engaging in argumentation so that we can determine what activities can be used in chemistry courses to help students learn to reason with evidence. We could then use these findings to develop evidence-informed activities and assessments to support students' argumentation skills. Chapter 2 describes our analysis of students' written arguments (N = 160) on a summative assessment item in organic chemistry related to the plausibility of an organic chemistry reaction. We analyzed students' arguments using an analytical framework with three dimensions: reasoning, granularity, and comparisons.

This was the first project I contributed to during my doctoral studies in the Flynn group. My primary intellectual contribution was helping conceptualize and analyze data using the comparisons dimension of the analytical framework, which emerged during discussions with Dr. Flynn and in response to reviewers' comments to early versions of the manuscript. I want to thank both Dr. Flynn and Nick Bodé, who led this project, for trusting and believing in me so early in my studies. Only with their guidance and support was I was able to meaningfully

contribute during a time in which I knew very little and had very little confidence in my own research skills and knowledge.

2.2 References for Introduction

European Union. (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. *Official Journal of the European Union*, L 394/19-L 394/18.

National Research Council. (2012). *A Framework for K-12 Science Education*. National Academies Press. https://www.nap.edu/read/13165/chapter/1

Organisation for Economic Cooperation and Development. (2006). *Assessing scientific, reading and mathematical literacy: a framework for PISA 2006.*

Social Sciences and Humanities Research Council. (2018). *Truth Under Fire in a Post-Fact World*. https://horizons.gc.ca/en/2018/10/19/the-next-generation-of-emerging-global-challenges/#truth-under-fire

2.3 Copyright

Reprinted from "Bodé, N. E.; Deng, J. M.; Flynn, A. B. Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *J. Chem. Educ.* **2019**, *96* (6), 1068–1082. doi.org/10.1021/acs.jchemed.8b00719" with permission from the American Chemical Society and Division of Chemical Education, Inc. Copyright © 2019 American Chemical Society and Division of Chemical Education, Inc.

JOURNAL OF CHEMICAL EDUCATION

Article

Cite This: J. Chem. Educ. 2019, 96, 1068−1082

pubs.acs.org/jchemeduc

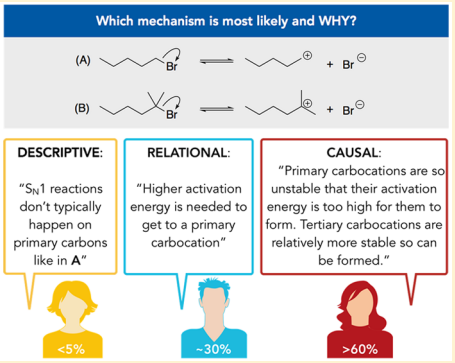
Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms

Nicholas E. Bodé, Jacky M. Deng, and Alison B. Flynn*

Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie Private, Ottawa, Ontario K1N 6N5, Canada

Supporting Information

ABSTRACT: Citizens in all roles need to be able to construct causal explanations of scientific phenomena and engage in coherent argumentation, supported by evidence and sound reasoning. These skills are needed in many contexts, from making everyday decisions to addressing complex global issues. One goal of instruction in this area is to get past assessment questions that can be answered using only memorized rules (e.g., S_N1 reactions are favored with tertiary halide substrates), toward answers that require reasoning with evidence (**why** is an S_N1 pathway favored under certain conditions?). Previous work has suggested that these reasoning skills may not be effectively taught, practiced, or learned in organic chemistry courses or directly assessed on exams. In this study, we explored students' ability to construct scientific arguments using evidence in the context of organic reaction mechanisms, specifically when comparing two proposed reaction mechanisms, making a claim as to which was most plausible, and providing reasoning. We used a qualitative coding scheme to analyze (1) the features and concepts students discussed in their arguments on exams, including the concept of granularity, (2) the connections (links) made between features and concepts, (3) the modes of reasoning revealed in the arguments, and (4) how the ideas and claims were explicitly compared. We found that ∼60% of arguments with correct claims established causal relationships between the relative stability of the carbocation intermediates and the relative activation energy barriers to the formation of these intermediates. However, the vast majority of students did not go to sufficient granularity to meet the expectations in the course, which they could have done by invoking Hammond's postulate and hyperconjugation in their arguments. Incorrect claims were often supported by causal arguments but these arguments were incorrectly based on drawing a causal relationship between the relative steric hindrance of the starting materials and activation energy. Over 60% of students provided a linear causal mechanistic argument to justify their claim, independent of whether their claims were correct. Over 90% of arguments at least partially compared key concepts explicitly, with over 20% doing so completely. The findings suggest that the majority of students in this study understood the need to provide cause-and-effect relationships to justify their answer, beyond identifying reaction features and their effects; however, students may have struggled to identify which features of those molecules were *relevant*.



KEYWORDS: *First-Year Undergraduate/General, Second-Year Undergraduate, Chemical Education Research, Organic Chemistry, Problem Solving/Decision Making, Reactive Intermediates, Mechanisms of Reactions, Carbocations, Testing/Assessment, Reactions*

FEATURE: Chemical Education Research

INTRODUCTION

We Need To Better Equip Learners To Reason Proficiently Using Evidence

In a world facing complex global issues that is increasingly being dominated by fake news and memes, citizens in any country need to develop proficiency in understanding and reasoning from evidence. These reasoning skills are needed in multiple contexts, such as making decisions to vaccinate, identifying and discarding fake news, and addressing global issues such as climate change. A core skill in such reasoning involves crafting scientific explanations and arguments. Researchers have cautioned that if we give priority to textbook correctness only, we lead students toward memorizing ideas rather than developing evidence-based reasoning skills.[1] Multiple calls have been made to provide coherent opportunities for students to build skills in constructing scientific explanations and arguments, including teaching approaches, learning activities, and assessments.[2−5]

However, research shows a challenge for learning as students' ideas about chemical causality often do not reflect consensus scientific models.[6,7] Moreover, students have few opportunities to gain scientific explanation and argumentation skills in

Received: September 3, 2018
Revised: April 25, 2019
Published: May 16, 2019

ACS Publications

© 2019 American Chemical Society and Division of Chemical Education, Inc.

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

Downloaded via UNIV OF OTTAWA on April 18, 2024 at 21:30:39 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

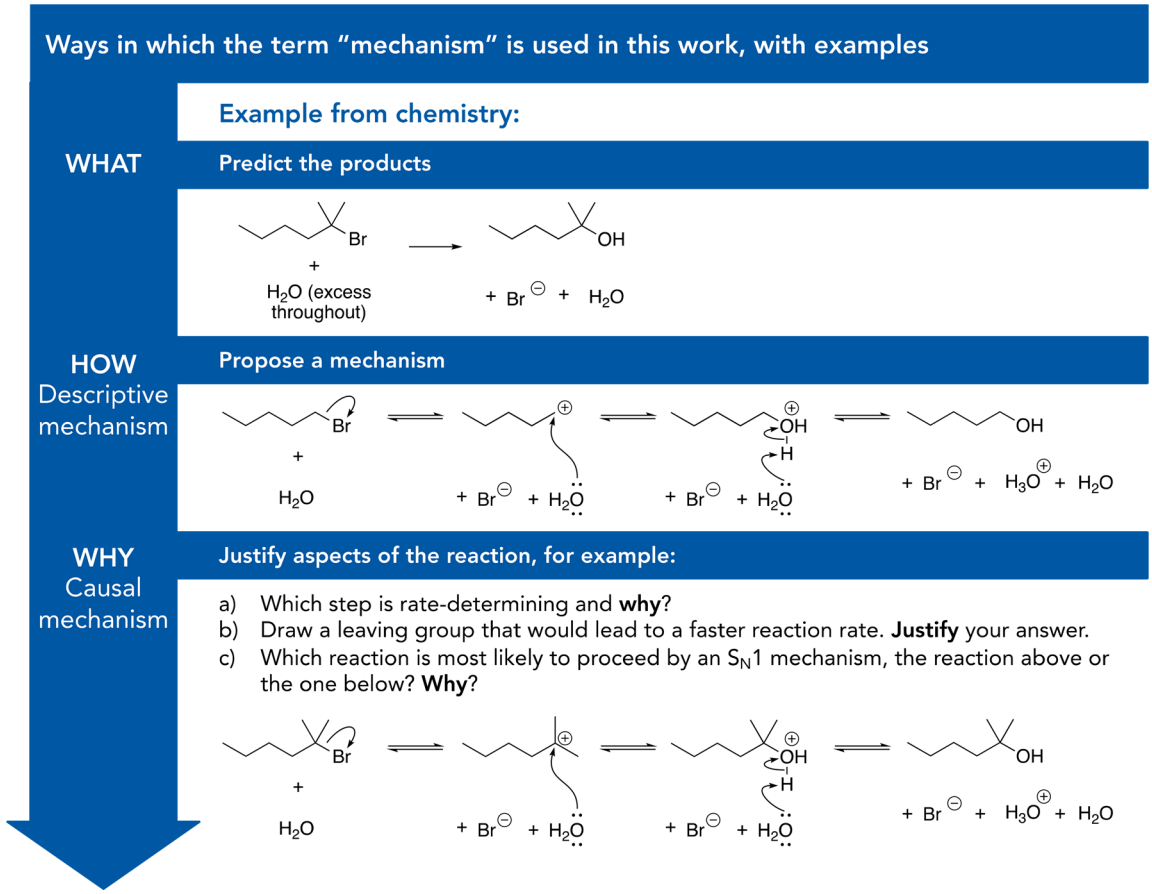


Figure 1. Questions in organic chemistry can address the what, how, and why of reaction mechanisms. Getting to the why engages causal mechanistic reasoning skills.

traditional course formats through standard learning activities assessments (formative and summative). For example, constructing scientific explanations appeared in less than 10% of American Chemical Society general chemistry exam items examined in 2016,[8,9] although it is a science practice emphasized in the National Research Council's Framework for K−12 education.[10] The vast majority of the assessment items on those exams did not include any science practices; of those that did, the majority required students to analyze and interpret data. An ACS Exam for organic chemistry did not assess students' ability to construct scientific explanations or arguments at all.[11]

Our goals in this study are to analyze students' reasoning and scientific arguments around specific organic chemistry reactions in a learning context that supports their skill development. Our intent is to use our findings to provide examples of assessment questions and associated skill-building activities that can be used in chemistry courses to help students learn to reason with evidence.

Causal Mechanisms

Chemistry reaction mechanisms comprise a major area in organic chemistry that provides an excellent basis for building scientific argumentation skills. The word "mechanism" has been defined in many ways,[12] including variations of the following: explanations that invoke entities having properties, interactions, activities, and organization that are responsible for the behaviors we observe.[13−15] Furthermore, the definition of "mechanism" has been discussed and used in various contexts.[2,14,16,17] Herein, the term "descriptive mechanism" refers to the elementary steps that comprise an overall reaction, or "how" the reaction occurs (Figure 1). The electron-pushing formalism (i.e., curved arrows) is a tool that helps describe the dynamic parts of the mechanism in which bonds are breaking and forming. The term "causal mechanism" refers to the reasons or cause behind a phenomenon or process, the "why". For example (Figure 1), chemistry questions can ask students to draw the products of a reaction (the "what"), to propose a mechanism (the "how" or descriptive mechanism), and to *justify* aspects of the reaction (the "why" or causal mechanism).[18,19] This description simplifies the literature on mechanistic reasoning (described more deeply in the Theoretical Framework section) but is meant as a tool to engage conversations, consider the questions we ask, and investigate possible lines of reasoning.[19] Getting to the "why" is key to engaging causal mechanistic reasoning skills.

Chemistry education efforts are being directed toward developing items (questions and activities) aligned with scientific argumentation skills and causal mechanistic reasoning in a variety of contexts. Previously, researchers have studied students' abilities and learning progressions about chemical causality.[17−23] A recent evaluation of the Organic Chemistry II curriculum at the authors' institution using a three-dimensional learning assessment protocol revealed that constructing scientific explanations and engaging in argumentation from evidence was represented in approximately 20% of course learning outcomes and 10% of midterm and final exam questions.[8,24] Since that time, new in-class learning activities and associated assessment questions to provide students with

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

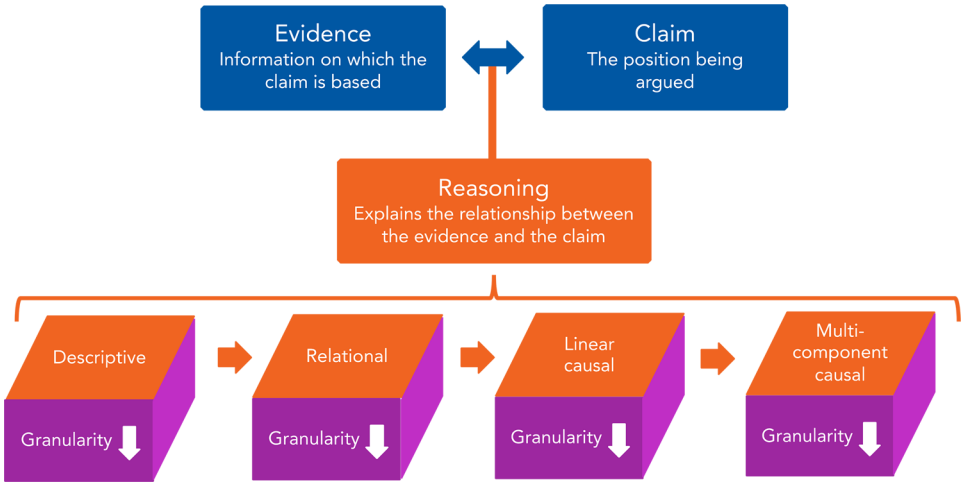
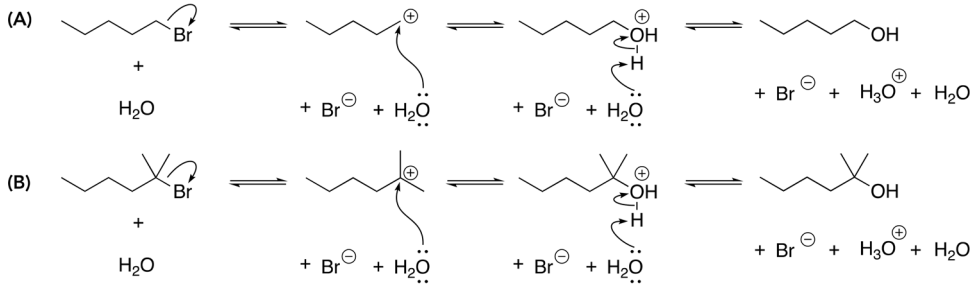


Figure 2. Overall theoretical framework: claim−evidence−reasoning, nature of evidence, granularity, modes of reasoning, and comparisons.

a. Draw and label the reaction coordinate diagram for the two reactions below that are shown proceeding by an S_N1 mechanism.



b. Which reaction is most likely to proceed by the mechanism shown? B

c. Justify your answer in part b, using your reaction coordinate diagram as part of your explanation.

Figure 3. Question analyzed for this study, taken from an Organic Chemistry II final exam.

opportunities to develop and demonstrate these practices have been piloted.

THEORETICAL FRAMEWORK

The theoretical framework for this study is composed of three components: the structure of the argument (claim−evidence−reasoning), the modes of reasoning, and the granularity of that reasoning (Figure 2).

Arguments Aim To Persuade Using a Claim, Evidence, and Reasoning

The terms explanation and argument have sometimes been used interchangeably in science contexts. Therefore, in the next paragraphs we describe the definitions that guide our work. Scientific explanations "explain observed relationships between variables and describe the mechanisms that support cause and effect inferences about them" (ref 1, p 67).[10] The goal of the explanation is to account for a consensually agreed fact or phenomenon to be explained.[25] Scientific explanation is more common than argumentation in classroom settings, where students are asked to reconstruct or identify explanations involving causal accounts of observable phenomena.[26]

In contrast, the goal of an argument is to persuade and justify a claim, and advancing that claim involves using data and reasoning. The claim is in doubt and must be advanced by constructing an argument about the fit between the known data and an explanation.[25] In other words, an argument is an assertion with a justification:[2728] "In science, reasoning and arguments are essential for identifying the strengths and weaknesses of a line of reasoning and for finding the best explanation for a natural phenomenon."[10] To use experimental data to support a theory, scientists must provide explicit links as to the relevance of the data to the claim (i.e., reasoning).[26] It should be noted that, in contexts involving discourse between individuals, argumentation is often constructed by multiple participants and can involve rebuttals.[26,29] We use the term *argument* herein to emphasize the definition that the goal of an argument is to use evidence to persuade and justify a claim.[25,26,29]

Scientific arguments have been analyzed using a number of different frameworks. In this study, one component of the framework analyzes the responses in terms of claim−evidence−reasoning (Figure 2),[27,30] an approach that has been used in other chemistry contexts, such as physical chemistry.[26] In part b of the question used in this study (Figure 3), students are asked to make a *claim* (the correct answer is "B"). They are then asked in part c to justify their answer. We analyzed their answers for the *evidence* used to support the claim, as well as their reasoning. The evidence used consisted of the features they identified from the question prompt (explicit and implicit) and concepts they leveraged in their arguments. We also analyzed the links made between those features and concepts. Other ways to consider evidence in science education,[14,15,31] and in chemistry more

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

Table 1. Comparison between Descriptive, Relational, Linear Causal, and Multicomponent Causal Arguments

Mode of Reasoning	Description	Example
Descriptive	Claims are supported solely by describing the properties of the reaction materials (starting materials, intermediates, products). Explicit (surface) features of the problem are described without further explanation.	*"B is more likely to proceed by the mechanism shown since S_N1 reactions don't typically happen on 1° carbons like in A"*
Relational	Explicit and implicit properties of the reaction materials are highlighted and discussed; comparisons may be made between the materials in each reaction presented. Relationships between properties of the reaction materials and their activities are discussed in a correlative fashion (i.e., a cause−effect relationship is not established).	*"The reaction B is more likely to occur than reaction A because the alpha carbon is a tertiary carbon. This means that it is more reactive and requires less energy for the reaction to occur. This means that the activation energy is lower and easier to reach. Also, an S_N1 reaction prefers a 3° carbon to occur. In contrast, reaction A contains a primary alpha carbon. This alpha carbon is not as reactive as an S_N1 reaction would need it to be and therefore it will require much more energy to be carried out. The activation energy will be higher and harder to react. The primary alpha is more suitable for an E2 or S_N2 reaction."*
Linear causal	All of the features of a relational argument are present and are accompanied by cause-and-effect relationships between the relevant properties of the reaction materials and their activities. Properties do not necessarily have to be relevant to a canonical explanation, but they must be relevant to the argument the student has provided.	*"An initial starting material of a 3° carbon next to the LG [leaving group] is favoured since it forms that 3° stable enough carbocation, a 1° carbon will not make a stable enough carbocation to form... a 1° carbon will not make a stable enough carbocation to form, so pathway A would have been a concerted SN2 mechanism with no carbocation formation step... the creation of the "A" carbocation is unlikely since it requires a lot more energy (higher activation energy) to form and thus is very unstable... reactions favour paths of less energy input (where it is easier to form products)"*
Multicomponent causal	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct inter- actions of several components. Causal stories are built. Explicit and implicit features of a system are noticed. Isolated: effects of several variables considered and weighed separately Integrated: arguments as interconnected stories of how different variables affect the entities involved	Examples not observed.

specifically,[2,6,17,23] have identified the entities and activities in a mechanism. Types of evidence can include features identified,[23,32] energetic and structural accounts,[23] static or dynamic approaches,[23,33] and dimensions with variations in explanatory power (i.e., chemical mechanism, causality).[34,35] *Reasoning* was analyzed in terms of modes of reasoning, as described below.

Modes of Reasoning

Reasoning can be analyzed from a number of different perspectives, and a number of applications have been applied in chemistry education research contexts, including modes of reasoning,[6,7] type I and II reasoning (including heuristics),[21,36−38] teleological reasoning,[17,38−40] abstractedness and abstraction,[41,42] and rules-, case-, and model-based reasoning.[43,44]

We analyzed students' arguments with modes of reasoning (i.e., descriptive, relational, linear causal, and multicomponent causal), which have also been called levels of complexity (Table 1, with additional examples in the Supporting Information).[6,7] *Descriptive* arguments contain statements of evidence, including surface features, without further information. *Relational* arguments discuss and connect implicit and explicit evidence, but in a correlative fashion absent of a cause−effect relationship (i.e., rules and connections are stated, without getting to the why). Causal arguments do describe the cause−effect relationships (i.e., *why* given evidence, properties, or concepts impact the effect, phenomenon, claim, etc.). In the first causal subtype, *linear causal*, single variables are connected directly to an effect. In the second causal subtype, *multicomponent causal*, multiple factors and variables in the evidence are weighed and considered, and then connected in a causal fashion (the *why*) together and to the claim. The mode of reasoning used will depend on the task, context, and expectations.[7,17,45] We used this theoretical framework because it is aligned with the intended learning outcomes and classroom activities in the course that relate to crafting scientific arguments.[10]

Granularity in Reasoning

At any level of complexity in reasoning, variations in granularity are possible, a concept that has also been described using various terms, including the following: depths of reasoning, sophistication, scales,[46] levels,[47] nested hierarchies,[31] emergence (with ideas of downward and upward causality),[48] and bottom-out reasoning.[13] For example, experts or students could be asked to explain how plants can have poisonous and nonpoisonous parts; an explanation may be provided at the population level (evolutionary explanation for how the differentiation arose), at the tissue/cellular level (getting toward cellular differentiation), at the molecular level (DNA's role), etc.[31] We consider the term granularity analogously to bottom-out reasoning. We did not expect any of the arguments in this study to bottom-out, and so the term granularity is found throughout this paper. We avoid the word sophistication because an answer could be sophisticated at multiple different levels of granularity, depending on the context and need. Each discipline has its understood need for levels of granularity.[13]

Experts and students alike may tailor the granularity of their scientific explanations or arguments to the context, purpose, and expectations. For example, if asked which reaction is most likely to proceed by an S_N1 mechanism and why, an expert or student might (correctly) respond by saying that reaction B is most likely *because* the substrate in B is a tertiary alkyl halide, while the

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

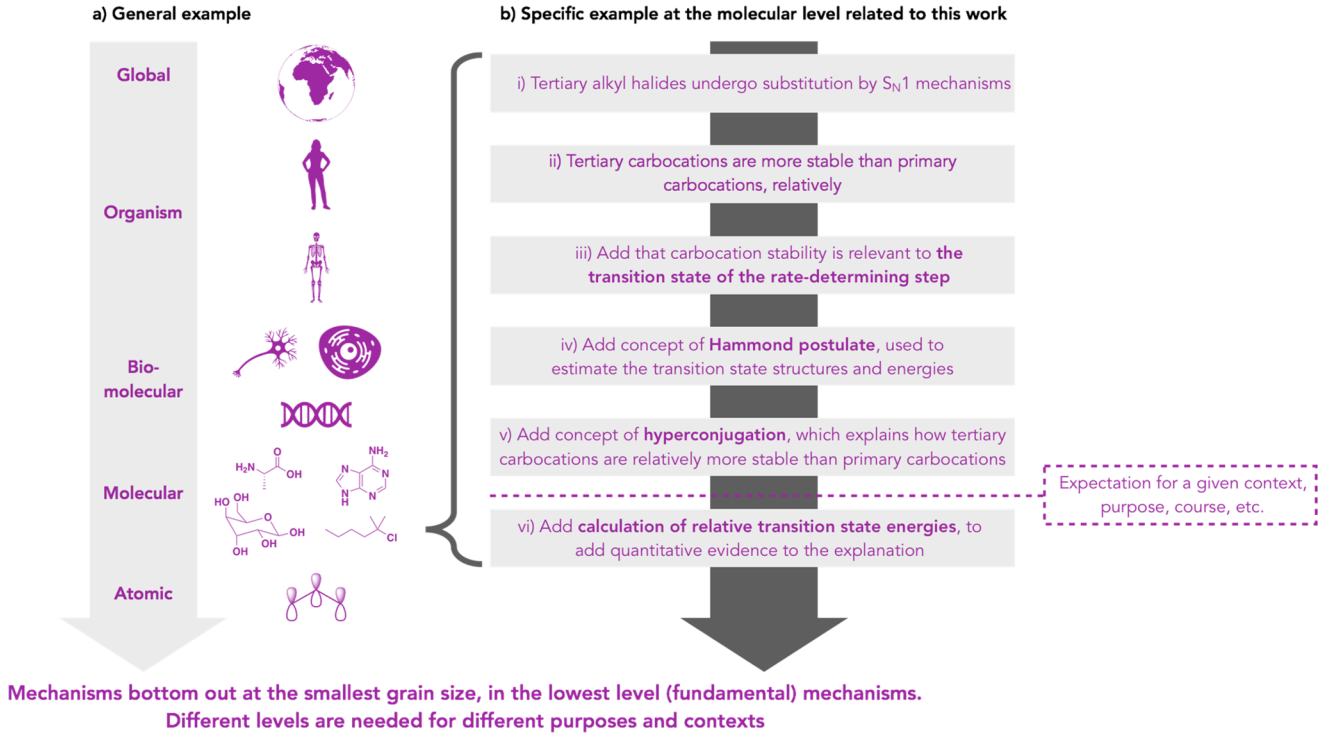


Figure 4. Variations in granularity are possible across and within levels in explanations or arguments that describe the same phenomenon: (a) a general example and (b) a specific example of levels of granularity.

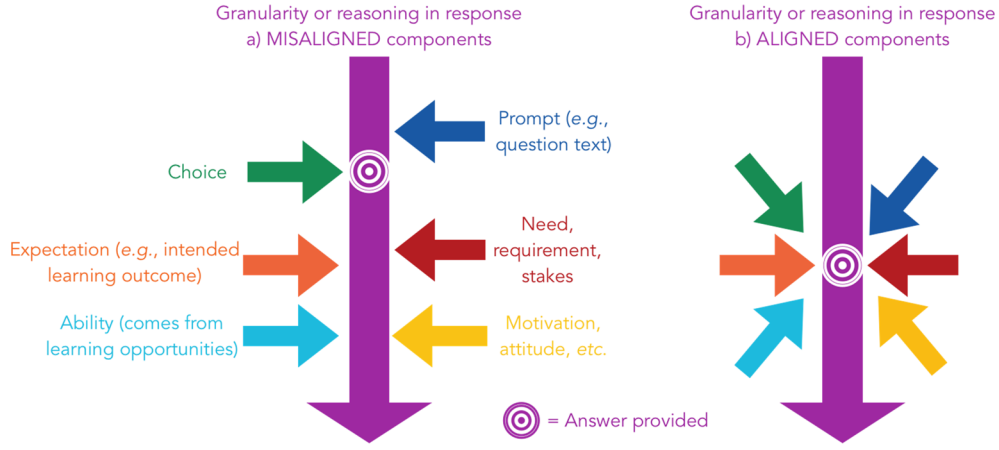


Figure 5. Granularity and mode of reasoning used in a scientific explanation or argument can depend on the prompt (question), expectation, need, choice, ability, motivation, etc., which can be (a) misaligned or (b) aligned.

substrate in A is a primary alkyl halide (Figure 4b,i. When asked for the underlying reason behind that statement, either could deepen their answer to include the relative stability of carbocations (Figure 4b,ii). If asked (or if expected) to describe the significance of relative carbocation stability to the mechanism, the respondent could include the idea that carbocation stability is relevant to the transition state of the rate-determining step (Figure 4b,iii). Since carbocations are not actually present in the transition state of the rate-determining step, the answer could become more sophisticated to include the relevance of carbocations to the transition state (Figure 4b,iv). Still, the answer could become more granular to include ideas of hyperconjugation (Figure 4b,v) and calculation of transition state energies (Figure 4b,vi) and go even deeper to include the molecular orbitals, physics, etc. Granularity can refer to moving across scales/levels (e.g., from molecular to atomic to electronic) or within the same level (deepening an explanation within the same scale).

While an expert may have the *ability* to continue expanding their explanation or argument to ever decreasing levels of granularity, the *actual* granularity of their scientific explanation or argument, even a multicomponent, causal one, will depend on their context and purpose. The granularity of an expert's explanation or argument can at some point even cross into another discipline where another expert's knowledge is needed, such as physics in this example. In an educational setting, students will similarly provide scientific explanations or arguments aligned with the context and purpose, *if* the expectations for a given context have been clearly communicated (e.g., activity, exam question) and the expectations are aligned with their abilities, motivation, etc. (Figure 5). Moreover, students need opportunities to gain skills in

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

constructing scientific explanations and arguments with a given level of granularity.

Careful consideration is needed in the wording of the questions (i.e., prompt) to elicit a particular level of granularity (or mode of reasoning); discussions of aspects of question construction can be found in the literature.[18] As Talanquer has pointed out, "Few learners, even those in advanced courses, seem to develop the ability to build explanations that mechanistically connect the electronic, atomic, molecular, multiparticle, and macroscopic scales [grain sizes]."[46] Moreover, when students engage in structure−property reasoning they tend to do it in associative manners, stating relationships but failing to properly justify them using mechanistic arguments.[46]

GOALS AND RESEARCH QUESTIONS

In the present study, we describe our analysis of students' responses to an exam assessment item designed to elicit causal mechanistic arguments when comparing mechanistic pathways. We wished to determine which features and concepts were included in the arguments, how the features and concepts contributed to a causal mechanism underlying the phenomenon, and how those concepts were connected. Specifically, we investigated the following research questions with respect to Organic Chemistry II students' responses on an exam question that asked students to decide which of two organic reaction mechanisms was most plausible (Figure 3):

(RQ1)
What features and concepts do students discuss and at what levels of granularity?

(RQ2)
What connections (links) are made between concepts?

(RQ3)
What modes of reasoning can be identified?

(RQ4)
How are ideas compared between related organic reaction mechanisms?

A substitution reaction was selected for the question as it is a central reaction in organic chemistry curricula and many student difficulties have been documented associated with that reaction.[49] Few organic chemistry assessments include questions that get to causal mechanistic arguments (*vide supra*), although the use of contrasting cases can elicit mechanistic reasoning.[23,36] Despite instruction, students may fail to understand the reasons behind the prevalence of one competing reaction pathway over another and may instead use heuristics, teleological explanations, and incorrect ideas to get the correct answer to the questions, as has been found in other contexts.[22,38] This study sought to examine students' arguments on an exam question that could not be fully answered with a set of memorized rules and instead required analyzing multiple competing reaction pathways (Figure 1, how) and the reasons for one mechanism/pathway to predominate in a given set of reaction conditions (Figure 1, why). In doing so, we hope to provide both example types that can elicit causal mechanistic thinking in organic chemistry and a framework for interpreting students' answers, whether for research or assessment.

Comparison between Reaction Pathways

The question prompt explicitly asked for a comparison between two possible reaction mechanisms. However, the previous aspects of the framework (i.e., claim−evidence−reasoning, modes of reasoning, and granularity) do not necessarily capture whether a comparison has explicitly been made between the two pathways. Without a comparison, a species cannot be more/less, bigger/smaller, or faster/slower. In this study, we identified whether the concepts in the argument were discussed in isolation from the other claim, partially compared with some but not all concepts in the argument discussed in relation to the other possible claim, and fully compared with all concepts in the argument (Table 2).

Table 2. Summary of Each Comparison Level with Description

Comparison Level	Description
Isolated	Concepts in argument of claim all discussed in isolation from the other possible claim. Concepts never used to compare/contrast between the claims.
Partially compared	Some (but not all) concepts in argument of claim discussed in relation to the other possible claim. These concepts are used to compare/contrast between the claims.
Compared	All concepts in argument of claim discussed in relation to the other possible claim. All concepts are used to compare/contrast between the claims.

METHODS

Setting and Course

This research was conducted in Organic Chemistry II (OC2) courses taught at a large, research-intensive Canadian university. OC2 is the students' second semester of organic chemistry. OC1 is offered in the winter semester of students' first year of their studies, and OC2 is offered in the summer and fall. Students can take these courses in English or French. OC2 is a 12-week course consisting of two weekly classes (200−400 students total, 1.5 h each, mandatory, lecture or flipped format)[50,51] and a tutorial session (1.5 h, optional, also called a recitation or discussion group). Assessment for the course is composed of two midterms, a final exam, online homework assignments, and class participation using a classroom response system. The course is composed of ∼75% Faculty of Science students, ∼17% Faculty of Health Sciences students, and ∼8% students from other faculties. The general topics covered in OC2 include reactions with σ electrophiles (e.g., S_N1/S_N2/E1/E2, and oxidation reactions), introduction to ^{1}H NMR and IR, additions to π electrophiles with leaving groups, and reactions with π nucleophiles (e.g., aldol reaction).[52,53]

Instructional Context

To investigate our research questions, we investigated students' ability to construct arguments about reactions and reaction coordinate diagrams (RCDs), specifically in the context of substitution reactions. Below, we list the intended learning outcomes that were provided to students for the module on substitution and elimination reactions that are pertinent to our research questions:

- Decide whether a reaction is likely to proceed via an E1/S_N1, E2, or S_N2 mechanism (i.e., descriptive mechanism)
- Draw and label the RCD for E1, S_N1, E2, and S^N2 reactions

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

- Explain a reaction mechanism and why the reaction is likely to proceed (or not) by that mechanism (i.e., causal mechanism)

This section will outline how these learning outcomes were taught, practiced, and assessed in the OC2 curriculum and how learning activities directly focused on constructing explanations and engaging in argumentation from evidence were incorporated into the curriculum.

Preclass. In the flipped classroom setting, students watch videos on the relevant background (posted for students on the course Web site) before coming to class, where time is primarily spent building on the preclass videos and answering questions using a classroom response system (Top Hat). In the curriculum, S_N1 and E1 reactions are taught together, as they both proceed through the formation of a carbocation intermediate, which is the rate-determining step of each reaction.

The first video in the course module on reactions with σ electrophiles focuses on E1 reactions; the video describes the electron-pushing mechanism for a generalized E1 reaction, an RCD for an E1 reaction, and the factors that influence the rate of E1 reactions (e.g., solvent polarity, leaving group stability, carbocation stability). Similar videos were also provided on the topics of S_N2 and E2 reactions.

The preclass materials in this module concluded with a video on differentiating E1, S_N1, and E2 using on a decision-making table, based on the factors listed above. The video does not always incorporate causal mechanisms for these factors; for instance, resonance stabilization of carbocations is discussed, but hyperconjugation is not.

In-Class. E1, S_N1, E2, and S_N2 reactions in OC2 are taught using Organic ChemWare animations as a learning tool,[54] in which an animation of the mechanism (with either arrows or molecular orbitals) is shown occurring in concert with the coordinate of the reaction displayed on an energy diagram at each point of the animation. Therefore, students have implicit opportunities to develop the ability to make connections between the corresponding elements of how organic reactions are represented using reaction mechanisms and how they are represented using RCDs. The development of these connections between analogous representations is referred to as coherence formation.[55]

In class, numerous activities asked students to make predictions/comparisons and justify their responses. The class analyzed a selection of responses both for correctness and completeness (i.e., presence of claim, evidence, and reasoning/justification, regardless of correctness). These interactive class activities occurred across a series of topics, including acid−base, substitution/elimination reactions, and carbonyl chemistry.

Postclass. Multiple problem sets addressing the intended learning outcomes in this module were provided for students to practice applying their knowledge to problems similar to what they would see on an exam. Answers to these problem sets were provided so that students could see specifically what was expected in their arguments of the relevant phenomena.

Midterm Exam. A similar question to the one analyzed for this study was given to students on their second midterm exam. A low success rate was observed on this problem, which we hypothesized was because students did not fully understand what the question was asking of them. The answer key to this midterm was provided to students so they could see the instructor's expectations explicitly, and the analogous problem on their final exam was reworded for clarity. Expectations were also discussed and practiced in class.

Data Source

To investigate our research questions, we analyzed students' responses to the final exam question shown in Figure 5, in which students were asked to draw reaction coordinate diagrams for two reactions that were both shown proceeding by an S_N1 mechanism. The second part of this question asked them to make a claim about which of these two reactions were more likely to proceed by this mechanism, and finally, they were asked to justify their responses. Only responses to parts a and c were analyzed for the purposes of this research. The authors' Institutional Review Board approved all stages of this project.

Expected Response

The following argument was expected in students' answers, on the basis of the intended learning outcomes for the course.[56]

> *Reaction B is more likely to proceed by an S_N1 mechanism than reaction A, based on the relative activation energies (E_a) of the rate-determining steps (RDS), which determine reaction mechanism feasibility. The E_a of the RDS is lower in pathway B, and this reaction will be more likely to occur via an S_N1 reaction. In contrast, the E_a of the RDS is likely insurmountable in pathway A, meaning this reaction could not proceed via the indicated mechanism.*
>
> *E_a is determined by the transition state structures' energy, which can be approximated using the Hammond postulate: the transition state bears the greater resemblance to the less stable species—reactant or product—of that elementary step.[57] The RDS is endergonic in these examples; therefore, the transition state in each reaction most closely resembles the carbocation intermediates formed.*
>
> *A primary carbocation is formed in pathway A, and a tertiary carbocation is formed in pathway B. Tertiary carbocations are better stabilized than primary carbocations (relatively) because the σ orbitals in the adjacent methyl groups' C−H bonds can overlap with and donate election density to the empty p orbital via hyperconjugation, stabilizing the p orbital. Primary carbocations are less stabilized by hyperconjugation because of fewer available overlapping σ orbitals, making their formation—and E_a leading to their formation—unlikely.*

The expected answer depends on the context, intended learning outcomes, purpose of the question, and other factors. A more sophisticated answer may have directly discussed the hyperconjugative stabilizing effects of the alkyl substituents on the forming p orbital as the leaving group departed and the carbon atom bearing that leaving group's hybridization changed, rather than establishing a relationship between the relative stability of the intermediates and transition state energy. An expert's answer might alternatively be less sophisticated than the expected one. Depending on the context in which they are being asked to provide the argument, *the most* sophisticated argument is not always necessary; less sophisticated arguments may be sufficient. In the course, regular communication and practice intend to clarify the expectations.

Coding Process

We identified the components of the expected answer that were consistent with the intended learning outcomes for the OC2 course and analyzed students' arguments for the presence of the components in the framework (aligned with each RQ). These steps established content validity for our initial coding scheme,

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

ensuring that we defined our initial coding scheme based on relevant structural features. Additionally, we defined the correctness and granularity of students' arguments based on canonical relationships between those structural features and the energy differences in the rate-determining step of each reaction. During the coding process, the coding scheme was refined to also include recurring concepts in students' arguments that were not present in our original coding scheme (i.e., not anticipated), even if they were associated with errors.

After identifying the features and concepts in the answers (RQ1), we identified the explicit links that were made between concepts and coded for whether those connections were correct (RQ2). Ideas were only coded if explicit meaning could be identified. For example, "lower in energy" and "more stable" were not coded together; although experts may make that link, we did not assume students would do so.

After describing the concepts students discussed and the nature of the connections between those concepts, we defined the modes of reasoning presented in students' arguments as descriptive, relational, linear causal, and multicomponent causal (Table 1, RQ3).[6,7]

The question used in this study required students to explain why one of two reactions was most likely to occur, which implicitly requires them to explain why the other is *not* likely to occur. Because of this comparison component, we also analyzed instances when students discussed similarities and differences in the two reactions while coding for what concepts students chose to discuss in their arguments about which reaction was more likely to occur via an S_N1 mechanism (RQ4). For example, if a student discussed activation energy, we also captured whether the student made an implicit or explicit comparison in the relative activation energy barriers for each reaction.

A more detailed account of the analysis framework is provided in the Supporting Information, which includes the list of codes with notes on their application, an explanation of the coding template, and a set of coding examples, in which students' full arguments are provided. In the SI, we also explain our analysis of these arguments in detail.

Inter-rater Reliability

A second rater analyzed a subset of students answers to establish inter-rater reliability.[58,59] The second rater used the codebook to code a different subset of 20 students' answers. Both raters then met to discuss how they coded each answer and reached 100% agreement on that subset of the data. Following this discussion, both raters agreed on revisions to the codebook to improve clarity. After the subset of 20 answers had been recoded and revisions to the codebook were made, both raters then coded another subset of 20 answers and repeated the process outlined above. This process was repeated four times, until 50% of the data set (80 answers) had been coded by both raters.

Inter-rater reliability was calculated using Krippendorff's α.[60] We calculated this statistic separately for each of the four categories of our analysis, summarized below:

1. The presence of specific concepts in students' answers
2. The presence of explicit links between concepts in students' answers
3. The presence of a comparison/contrast between reactions A and B with respect to each concept we identified in students' answers
4. The sophistication of students' overall arguments (descriptive, relational, or causal)

For the first three iterations of this procedure, we were unable to reach the threshold of acceptability for inter-rater reliability (Krippendorff's α = 0.67) for all four of these categories. However, after the fourth round of individually coding 20 exams, we obtained Krippendorff's α values that exceeded this threshold for each aspect of the coding scheme. Values are provided in Table 3. We then reached consensus on any remaining disagreements within that set of 20 exams and revised the codebook as required, and the first author recoded the remaining half of the data set using the revised codebook.

Table 3. Inter-rater Reliability Values Obtained for Each Coding Category Used in This Study

Analysis Category	Krippendorff's α[a]
Presence of specific concepts	0.86
Presence of explicit links between concepts	0.75
Presence of a comparison/contrast between reactions A and B with respect to each concept	0.77
Sophistication of students' overall arguments	0.90

[a]Acceptable agreement = 0.67; definitive agreement = 0.80.

Visualization of Concept Networks

We visualized the connections that students made between concepts in their answers using Gephi software (v 0.9.2). In these visualizations, nodes represent the concepts that we identified during the coding process. The thickness of a given edge (i.e., line between two nodes) corresponds to the number of times a student made an explicit connection between the concepts connected by that edge. In other words, the thickest edges correspond to the concepts that were connected most frequently. A node without edges indicates that the concept represented by that node was not linked to any other concepts by any student. Using a ForceAtlas algorithm, Gephi placed concepts that were most frequently connected by students to other concepts centrally within the node network. Tight clusters of nodes indicate groups of concepts that were frequently discussed together. Further details about data treatment in the Gephi visualizations can be found in the SI.

The relative distance between nodes is meaningful within an individual Gephi visualization, but this distance cannot be meaningfully compared *between* visualizations. In each individual network visualization, the networks had to be expanded to different extents to avoid overlapping nodes. Additionally, nodes that were not connected at all (i.e., lacked edges) were placed manually after running the algorithm; the distance of these unconnected nodes from the other nodes is therefore not meaningful.

RESULTS AND DISCUSSION

RQ1: What Features and Concepts Are Discussed in the Arguments and at What Levels of Granularity?

Most students (73%) correctly identified the reaction in Figure 5 that would more likely proceed via an S_N1 mechanism. In answers with a correct claim, the most frequently discussed concepts were stability, carbocation substitution, and activation energy (Figure 6). These three concepts were all relevant to the claim being made, and the vast majority of students who discussed these concepts did so without errors. However, the expected answer also included concepts of rate-determining step and transition state. Each of these concepts was discussed in less than 20% of arguments. Additionally, few arguments explicitly

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

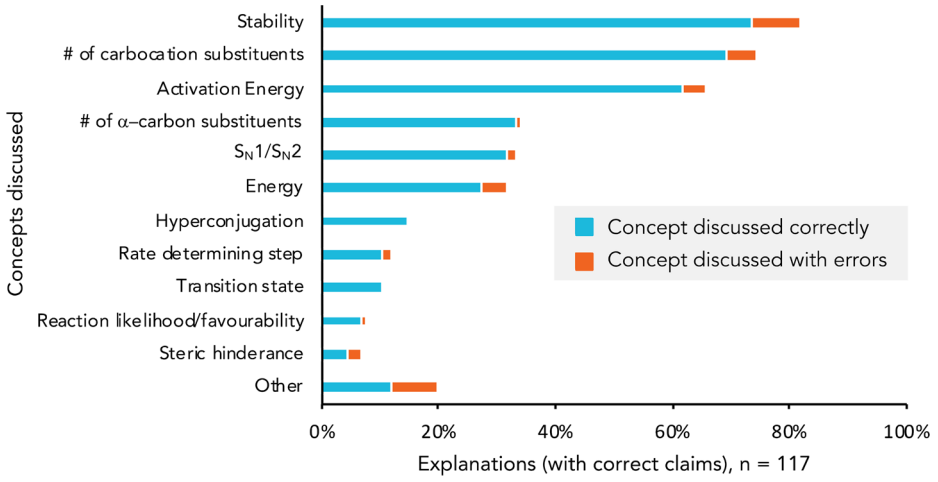


Figure 6. Concepts discussed in answers with the *correct* claim that reaction B is more likely to proceed via an S_N1 mechanism. $N = 117$.

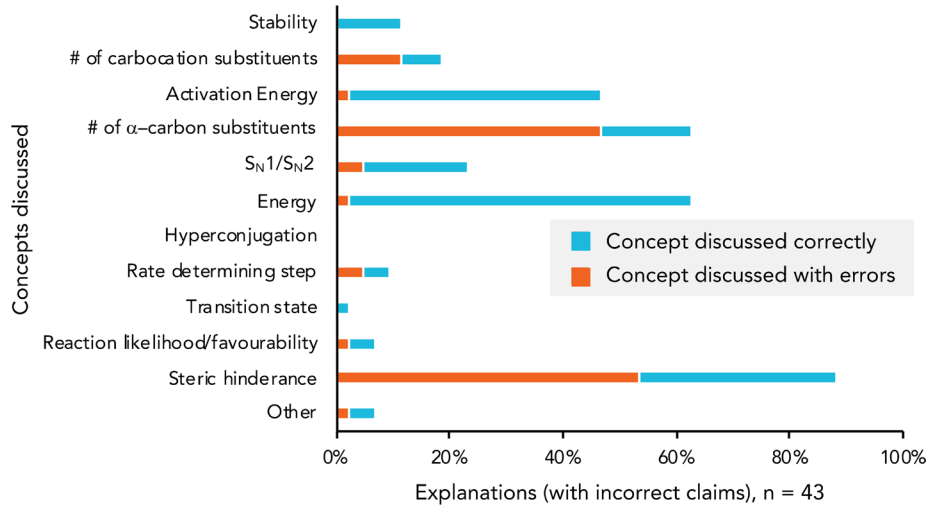


Figure 7. Concepts discussed in answers with the *incorrect* claim that reaction A is more likely to proceed via an S_N1 mechanism. $N = 43$.

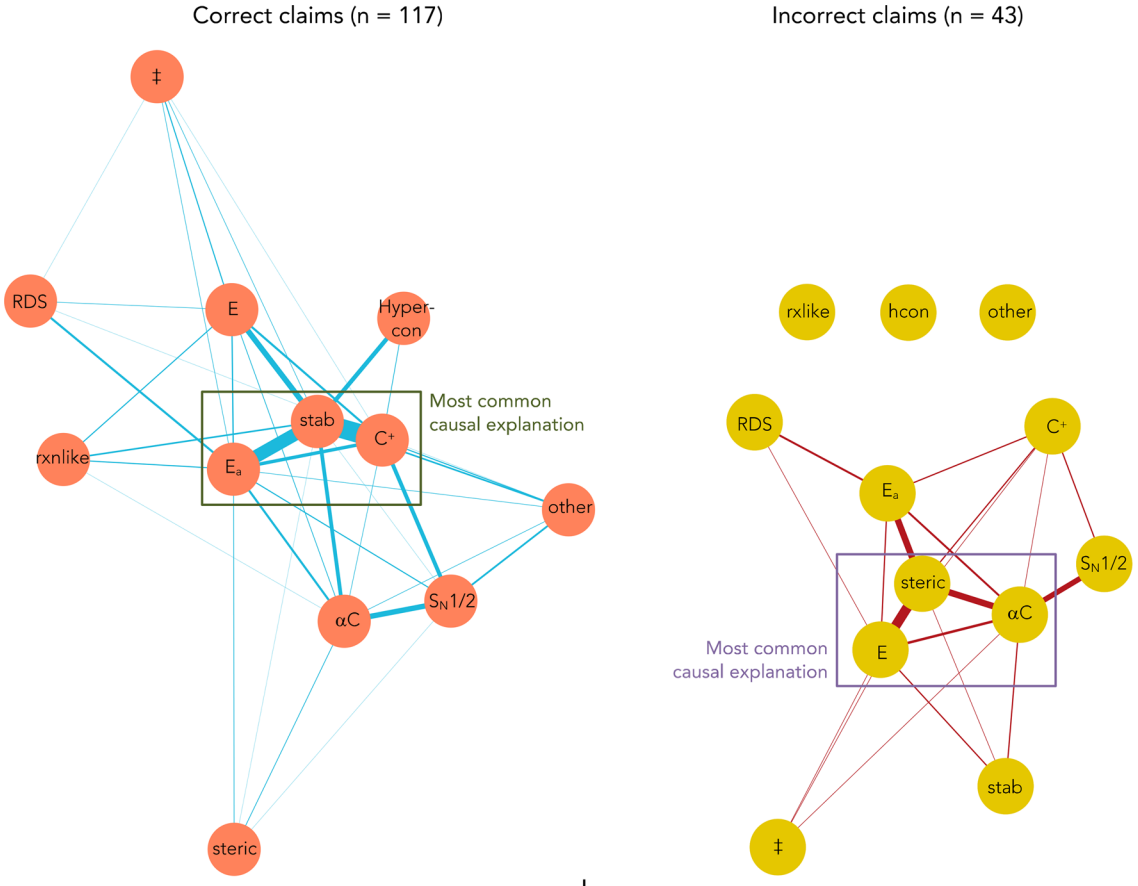
incorporated concepts of hyperconjugation or Hammond's postulate or provided other discussion of why relative carbocation stability is relevant to the transition state structure and energy. This level of granularity was part of activities in the course—including arguments—and was expected in the exam; however, the exam question did not give specific direction as to what level of granularity was required.

For arguments with incorrect claims, the most commonly discussed concepts were steric hindrance, α carbon substitution, energy, and activation energy. The latter two ideas were almost always discussed with errors (Figure 7). Concepts of steric hindrance and α carbon substitution were frequently discussed without errors; in those cases, students generally explained that tertiary α carbons are more sterically hindered than primary α carbons. The steric factors are less relevant to a claim about the kinetics of an S_N1 reaction; this evidence would be relevant if part of the argument involved explaining why the reaction would *not* go through an S_N2 mechanism. Identifying relevant features (explicit or implicit) is a core component of generating a scientific argument, as the argument depends on the quality of the evidence invoked. Students may not have recognized relevant features (e.g., carbocation, transition state, etc.) or may have fragmented knowledge that they are struggling to integrate, similar to researchers' findings in other contexts.[3,61] Few arguments invoked hyperconjugation to explain why the tertiary carbocation is more stable than the primary; in other words, few arguments went down to the expected level of granularity.

RQ2: What Connections (Links) Are Made between Concepts?

Connections in Arguments with a *Correct* Claim. The three most discussed concepts (stability, carbocation substitution, and activation energy) were also the most connected concepts (Figure 8, left). Generally, these arguments first discussed carbocation substitution, comparing the primary carbocation intermediate in reaction A to the tertiary carbocation intermediate in reaction B. Next, students explained that the tertiary carbocation is more stable than the primary carbocation, and therefore, the activation energy barrier to its formation is lower. We considered these three-component arguments to be causal arguments, but the expected answer was more granular. Most students who provided correct claims did not discuss the link between the relative energy levels of the intermediates and the activation energy (e.g., using Hammond's postulate). Common incorrectly identified features (e.g., steric hindrance) could be used as the basis for class activities or problem sets to help students analyze the features present in reactions and build skill deciding on the relevant features.

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082



Abbreviation	Code	Abbreviation	Code
αC	# of α-carbon substituents	‡	Transition state
C+	# of carbocation substituents	E	Energy
RDS	Rate determining step	$S_N1/2$	S_N1/S_N2
stab	Stability	steric	Steric hinderance
E_a	Activation energy	Hyper-con	Hyperconjugation
rxnlike	Reaction likelihood/favourability	other	Other

Figure 8. Network visualizations of the total number of connections made between concepts in arguments for correct (left) and incorrect (right) claims. The edge thicknesses represent the percentage of arguments that made a connection between the linked concepts.

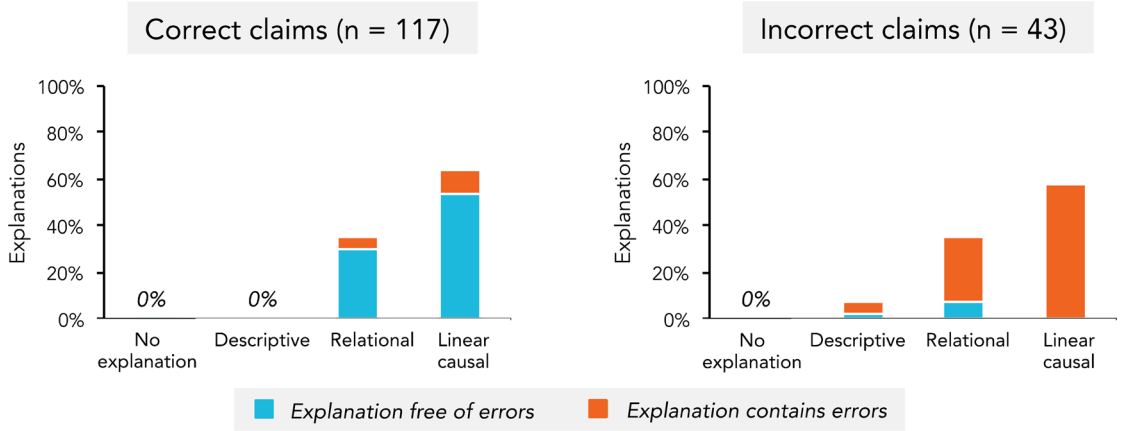


Figure 9. Modes of reasoning found in students' arguments to support their claim for whether reaction A or B (Figure 3) would be more likely to proceed via an S_N1 mechanism. $N = 160$.

Connections in Arguments with an *Incorrect* Claim. Far fewer links were observed among concepts in arguments for incorrect claims, but the most common causal argument was centered on the idea of steric hindrance, which was incorrectly

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

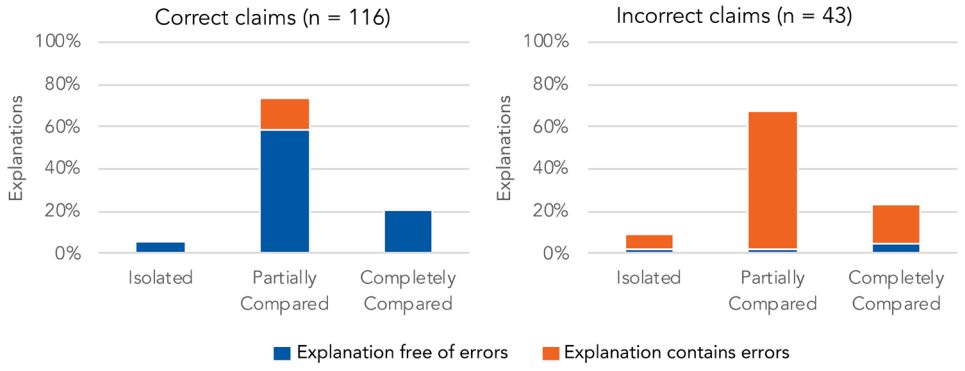


Figure 10. Distribution of comparison levels (regardless of level of reasoning) in students' arguments.

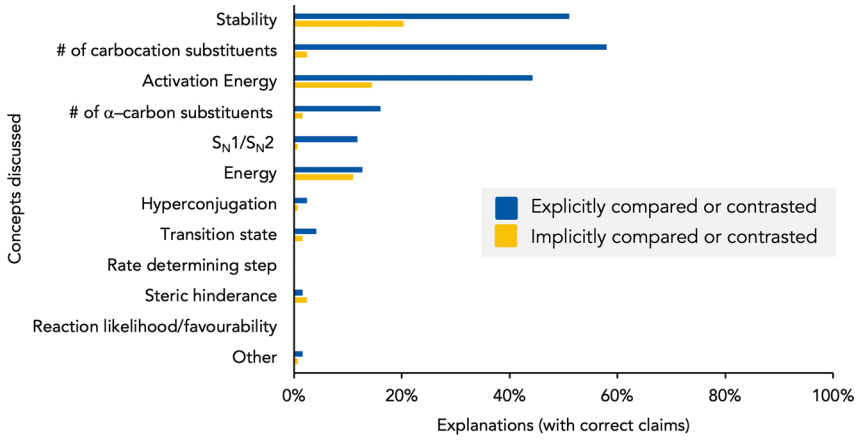


Figure 11. Concepts for which comparisons were made between reactions A and B. The data shown are for answers with the *correct* claim that reaction B is more likely to proceed via an S_N1 mechanism. $N = 117$.

used to justify why the activation energy barrier to the formation of the product in reaction B would be higher (Figure 8, right). This reasoning was more consistent with trying to rationalize the outcome of a reaction proceeding via an S_N2 mechanism, despite the question explicitly having stated and shown the reactions proceeding via S_N1 mechanisms. Because we lack detailed knowledge of students' mental models of these reaction mechanisms, we do not know why a steric argument was invoked so frequently.

RQ3: What Modes of Reasoning Can Be Identified in Students' Arguments?

Relational arguments were less common but generally took one of two forms. The first common relational argument discussed α carbon substitution and S_N1/S_N2. These arguments were rule-based in nature, generally explaining that tertiary alkyl halides react via an S_N1 mechanism while primary alkyl halides react via an S_N2 mechanism, without giving a reason why. The other common relational argument discussed carbocation substitution and either energy or activation energy, generally explaining that it takes less energy to form a tertiary carbocation than a primary carbocation, without invoking stability as a cause.

Most students (63%) provided linear causal arguments to support their claim (Figure 9, left), regardless of whether the claim itself was correct; multicomponent causal arguments were not found. The proportion of causal arguments was not significantly different between arguments with correct and incorrect claims, $\chi^2(1) = 0.477$, $p = 0.581$, with a small effect size, $\phi = 0.05$.

Correct claims were almost exclusively (98%) associated with causal or relational arguments. In contrast, incorrect claims contained more descriptive arguments (7%). A small proportion of incorrect claims were supported by error-free arguments, mainly because these arguments did not go into enough detail for an error to be made.

We had expected to find more correct claims justified incorrectly using irrelevant evidence (e.g., sterics). Although most of the correct claims contained correct evidence in this study, we anticipate finding instances of a correct claim supported by incorrect evidence (in future work, for example). For example, students might have used a steric argument to explain why the S_N1 mechanism is most likely in reaction B; this approach has been referred to as the application of one or more non-normative concepts.[62] Even the use of non-normative concepts can be useful for instruction, as they can provide a foundation for further instruction.

We are encouraged by the high proportion of higher modes of reasoning (60% linear causal with the majority of the rest being relational). Other research has reported lower proportions or no instances of causal mechanisms;[22] some studies, particularly in redesigned curricula, have found similar levels of causal mechanistic arguments.[18,40] However, research in organic chemistry settings found that participants who constructed arguments with higher complexity often involved follow-up questions from the interviewer.[23] That finding suggests that scaffolding in educational design can help students learn to construct more advanced arguments, analogous to the role of the interviewer. Our previous work has identified instances of causal

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

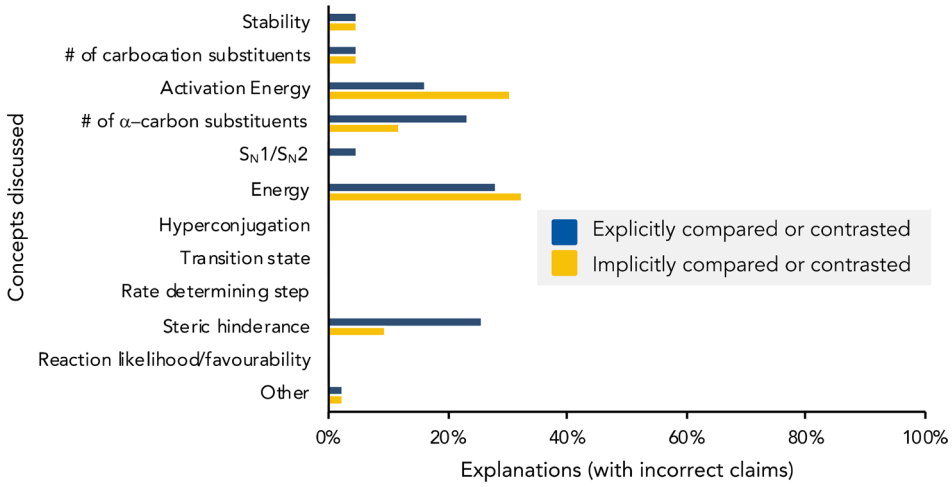


Figure 12. Concepts for which students made a comparison or contrast between reactions A and B. The data shown are for answers with the *incorrect* claim that reaction A is more likely to proceed via an S_N1 mechanism. $N = 43$.

reasoning, even though the research tasks (i.e., interview prompts) were not designed to explore that type of reasoning.[32] Students' instincts to seek causal factors provide another excellent foundation for further instruction.

We do not think that we should be aiming for eliciting multicomponent causal reasoning for every question type that requires an argument or explanation. Similar to adapting granularity of an argument to a given context or purpose, various modes of reasoning can be appropriate for different purposes: for example, "Expert chemists are known to rely on simple correlations and causal links to make sense of chemical processes, particularly when thinking about well-known classes of reactions."[7] Graduate student participants in that study used a variety of modes of reasoning, which may suggest high adaptability to different situations.[7] An important goal in an education setting, then, is for students to be able to identify the mode of reasoning used in an argument and to construct an argument with a given mode of reasoning (or level of complexity).

RQ4: How Are Ideas Compared Concepts between Related Organic Reaction Mechanisms?

The majority of arguments partially compared the features in one claim to another possible claim (Figure 10), regardless of whether the claim was correct or incorrect, $\chi^2(1) = 0.908$, $p = 0.461$. The mode of reasoning was similarly not a factor in the completeness of the comparison (see SI for details). Similar to the modes of reasoning analysis, students in this study displayed the same level of skill in making explicit comparisons whether their claim was correct or incorrect.

Figure 11 (correct claims) and Figure 12 (incorrect claims) show the frequency at which arguments contained comparisons between the two reactions with respect to each concept that they discussed in their arguments. In the problem we analyzed, the most relevant structural feature was the relative number of substituents on the C atom bearing the leaving group in each reaction (or on the carbocation intermediate). In arguments of correct claims, an explicit comparison was made between one of these structural features in 72% of arguments, while only 28% of arguments with incorrect claims contained an explicit comparison of structural features. This difference between arguments of correct and incorrect claims was significant, $\chi^2(1) = 25.97$, $p < 0.001$, with a medium effect size, Cramer's $V = 0.40$.

Similar trends were observed with respect to the energetics related to these structural features. The most commonly discussed energetic concepts in arguments of correct claims were stability and activation energy; these were compared between reactions in 51% and 44% (respectively) of these arguments. In contrast, the most commonly discussed energetic concepts in arguments of incorrect claims were steric hindrance and energy, which were only compared between reactions in 25% and 28% (respectively) of these arguments.

The preceding results indicate that although a majority of students in this setting were able to provide a causal mechanism in their argument (independent of the correctness of their claim), not making explicit comparisons between structural features may have been a limiting factor in their ability to make a correct claim about the problem they were presented. Since we observed that most students tried to make explicit links between structural and energetic features of the reaction that they believed were relevant, the limitation in their arguments appears to primarily be the lack of explicit comparisons of these structural and energetic features between the two reactions they were asked to reason about.

LIMITATIONS

This study only focused on students' ability to provide causal arguments about unimolecular substitution reactions; findings are not generalizable to other domains of organic chemistry because students' ability to provide causal arguments about scientific phenomena may be dependent on the sophistication of their content knowledge. In this setting, the majority of students provided causal arguments regardless of whether their claim was correct. Of the students who did not provide causal arguments, we do not know if they were unable to provide a causal argument for their claim, if they needed further scaffolding to do so, or if they did not know they were required to do so (i.e., the question did not make this requirement explicit). We cannot make claims of transferability of the findings to other contexts because the findings were dependent on the instructional context (e.g., flipped classroom, specific learning activities, students' knowledge of and experience with assessments).

The preclass instructional video in our setting that focused *specifically* on differentiating E1, S_N1, and E2 was based on a decision-making table, which likely played a role in some students seeming to develop their understanding of these

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

reactions based on the rules presented in this table. Although the causal mechanisms underlying these rules were provided to students in other learning contexts, evidence suggests that people construct mental models that are "good enough" rather than optimal.[63] Therefore, students might not have seen the need to understand the causal mechanisms that explain differences in stability of intermediates and activation energy barriers for competing mechanisms if they could rationalize the outcomes without that understanding. Future improvements to the curriculum might involve revising preclass instruction to more explicitly emphasize the need to provide the causal mechanisms underlying these rules. In other words, the mode of reasoning and level of granularity required need to be well-communicated.

Finally, students used the reaction coordinate diagram (RCD) drawings they provided in part a of the question as evidence in their arguments. Although a causal argument could be made without an RCD of why reaction B is more likely to proceed via S_N1 than A, an incorrect RCD may have subsequently impacted their answer.[64,65]

CONCLUSIONS

The majority of students in this setting provided causal arguments for the claims they were making, regardless of whether that claim was correct or incorrect. The correct claim was typically supported by the fundamentally necessary evidence that would explain the predominance of an S_N1 reaction mechanism. However, most responses did not include the expected granularity, specifically that relative carbocation stability would be justified by hyperconjugation and that Hammond's postulate would be invoked to explain the relationship between the relative stability of the carbocation intermediates in each pathway and the activation energy barrier to their formation.

Meanwhile, incorrect claims were still supported by a linear causal mechanistic argument, although the most common argument provided for these claims incorrectly used steric reasons as evidence. Difficulty using the appropriate evidence has been identified in other contexts.[66] Additionally, we observed fewer links between concepts in answers that had incorrect claims. The primary difference between arguments of correct and incorrect claims was the frequency at which students explicitly compared structural and energetic features between the two reactions that were shown. Almost all arguments contained explicit comparisons between the reaction mechanisms being considered, although only approximately 20% of arguments had complete comparisons of all features and concepts between reactions.

This work provides evidence in one context of how students construct scientific arguments in a high stakes summative assessment (i.e., final exam), the aspects of the reaction and question prompts they deem to be relevant, and their skill in connecting ideas together. Assessment items such as these, used in formative (e.g., in class) and summative (e.g., exams) settings, provide opportunities for students to develop scientific argumentation skills.[10]

IMPLICATIONS FOR LEARNING, TEACHING, AND RESEARCH

One way to make expectations clear for students involves defining the expectations using descriptive intended learning outcomes, teaching students those concepts, giving students the opportunity to practice in and out of class with feedback (formative and summative assessments), and ensuring assessments are aligned with those learning outcomes and teaching/learning activities. Opportunities to practice can include constructive (e.g., explaining concepts from a video, comparing and contrasting with prior knowledge) and interactive tasks (e.g., debating with a peer about justifications; discussing similarities and differences).[7,67] In-class activities provide information to learners and educators about how an activity's instructions are being interpreted, allowing for alignment of expectations and understanding prior to higher stakes summative assessments.

Although we did not look for heuristic reasoning in our analyses, this type of reasoning and related approaches (e.g., rules-based, type I) are likely to impact students' arguments as has been found in other contexts.[36,37,43] Heuristics and other shortcuts in reasoning are useful in many situations to help us avoid cognitive overload and simplify mechanistic arguments.[40] Helping students learn about when and how to move past these approaches would be valuable.[37] Similarly, we did not analyze the data for teleological arguments, but such strengths and limitations of teleological arguments can be considered in many ways analogously to heuristics: useful constructs when used mindfully and appropriately, but barriers to deep understanding and scientific reasoning when used incorrectly or in inappropriate situations.[40] An alternative curriculum that fosters structure−property reasoning indicated that students benefit from consistent and continued opportunities to analyze data, identify patterns, build and apply models, engage in argumentation with peers, generate and communicate arguments, etc.[46]

If students are to achieve higher modes of reasoning and to develop an ability for multivariate problem-solving, we need to align our educational approaches to support their success, including providing sufficient time for students to develop these skills,[68] carefully designing educational opportunities to support students' success, and assessing for the learning outcomes that we wish students to achieve. In organic chemistry, examples for eliciting causal mechanistic reasoning can involve comparing reaction pathways. Moreover, for students to transfer their abilities to new situations, they have to be able to identify their own reasoning skills and approaches.[68] Integrating metacognitive skill-building opportunities into courses and textbooks offers ways for students to build these skills.[48,69]

Because scientific argumentation skills are meant to persuade using evidence and reasoning, they can be extended beyond a given course to broader contexts.[29,70] Helping students develop persuasion skills will not only involve the cognitive and metacognitive aspects discussed herein, but also require addressing social aspects.[70] Students will benefit from persuasion skills in a variety of contexts, including convincing friends, colleagues, and family about key issues such as vaccination and climate change.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.8b00719.

Codebook used in our analyses, including the following: list of codes with notes on their application, the coding template, a set of coding examples in which students' full

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

arguments are provided, and a more detailed analysis of comparisons of reactions in the arguments (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Alison.Flynn@uOttawa.ca.

ORCID

Alison B. Flynn: 0000-0002-9240-1287

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Keith Lapierre for his work on the inter-rater analysis.

REFERENCES

(1) Russ, R. S.; Coffey, J. E.; Hammer, D.; Hutchison, P. Making Classroom Assessment More Accountable to Scientific Reasoning: A Case for Attending to Mechanistic Thinking. *Sci. Educ.* **2009**, *93*, 875−891.

(2) Talanquer, V. Importance of Understanding Fundamental Chemical Mechanisms. *J. Chem. Educ.* **2018**, *95*, 1905.

(3) Cooper, M. M. Why Ask Why? *J. Chem. Educ.* **2015**, *92* (8), 1273−1279.

(4) Cooper, M. M.; Caballero, M. D.; Ebert-May, D.; Fata-Hartley, C. L.; Jardeleza, S. E.; Krajcik, J. S.; Laverty, J. T.; Matz, R. L.; Posey, L. A.; Underwood, S. M. Challenge Faculty to Transform STEM Learning. *Science (Washington, DC, U. S.)* **2015**, *350* (6258), 281−282.

(5) Cooper, M. M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116−1122.

(6) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 10−23.

(7) Weinrich, M.; Talanquer, V. Mapping Students' Modes of Reasoning When Thinking About Chemical Reactions Used to Make a Desired Product. *Chem. Educ. Res. Pract.* **2016**, *17*, 394−406.

(8) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11* (9), No. e0162333.

(9) Reed, J. J.; Brandriet, A. R.; Holme, T. A. Analyzing the Role of Science Practices in ACS Exam Items. *J. Chem. Educ.* **2017**, *94* (1), 3−10.

(10) National Research Council (NRC). *A Framework for K-12 Science Education*; The National Academies Press: Washington, DC, 2012.

(11) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.

(12) Bhattacharyya, G. From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism. *J. Chem. Educ.* **2013**, *90* (10), 1282−1289.

(13) Machamer, P.; Darden, L.; Craver, C. F. Thinking about Mechanisms. *Philos. Sci.* **2000**, *67* (1), 1−25.

(14) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Sci. Educ.* **2008**, *92* (3), 499−525.

(15) Darden, L. Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. *Philos. Sci.* **2002**, *69* (S3), S354−S365.

(16) Tabery, J. G. Synthesizing Activities and Interactions in the Concept of a Mechanism. *Philos. Sci.* **2004**, *71* (1), 1−15.

(17) Caspari, I.; Weinrich, M.; Sevian, H.; Graulich, N. This Mechanistic Step Is "Productive": Organic Chemistry Students' Backward-Oriented Reasoning. *Chem. Educ. Res. Pract.* **2018**, *19*, 42−59.

(18) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid−Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703−1712.

(19) Braaten, M.; Windschitl, M. Working toward a Stronger Conceptualization of Scientific Explanation for Science Education. *Sci. Educ.* **2011**, *95* (4), 639−669.

(20) Cole, R.; Becker, N.; Towns, M.; Sweeney, G.; Wawro, M.; Rasmussen, C. Adapting a Methodology from Mathematics Eduation Research to Chemistry Education Research: Documenting Collective Activity. *Int. J. Sci. Math. Educ.* **2012**, *10* (1), 193−211.

(21) McClary, L.; Talanquer, V. Heuristic Reasoning in Chemistry: Making Decisions about Acid Strength. *Int. J. Sci. Educ.* **2011**, *33* (10), 1433−1454.

(22) Becker, N.; Noyes, K.; Cooper, M. M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, *93* (10), 1713−1724.

(23) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chem. Educ. Res. Pract.* **2018**, *19*, 1117.

(24) Raycroft, M.; Flynn, A. B. Next Steps Toward Improving the Organic Chemistry Curriculum at the University of Ottawa. In *Canadian Society for Chemistry: 100th Canadian Chemistry Conference and Exhibition*; Toronto, ON, 2017.

(25) Osborne, J. F.; Patterson, A. Scientific Argument and Explanation: A Necessary Distinction? *Sci. Educ.* **2011**, *95* (4), 627−638.

(26) Becker, N.; Rasmussen, C.; Sweeney, G.; Wawro, M.; Towns, M.; Cole, R. Reasoning Using Particulate Nature of Matter: An Example of a Sociochemical Norm in a University-Level Physical Chemistry Class. *Chem. Educ. Res. Pract.* **2013**, *14* (1), 81−94.

(27) McNeill, K. L.; Lizotte, D. J.; Krajcik, J.; Marx, R. W. Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *J. Learn. Sci.* **2006**, *15* (2), 153−191.

(28) Kuhn, D. *The Skills of Argument*; Cambridge University Press: New York, 2011.

(29) Toulmin, S. *The Uses Of Argument*; Cambridge University Press: Cambridge, 1958.

(30) McNeill, K. L.; Krajcik, J. S. Middle School Students' Use of Appropriate and Inappropriate Evidence in Writing Scientific Explanations. In *Thinking with Data: The Proceedings of the 33rd Carnegie Symposium on Cognition*; Lovett, M. C., Shah, P., Eds.; Erlbaum: Mahwah, NJ, 2007; pp 233−265.

(31) Southard, K. M.; Espindola, M. R.; Zaepfel, S. D.; Bolger, M. S. Generative Mechanistic Explanation Building in Undergraduate Molecular and Cellular Biology. *Int. J. Sci. Educ.* **2017**, *39* (13), 1795−1829.

(32) Webber, D. M.; Flynn, A. B. How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *J. Chem. Educ.* **2018**, *95* (9), 1451−1467.

(33) Bongers, A.; Northoff, G.; Flynn, A. B. Working with Mental Models to Learn and Visualize a New Reaction Mechanism. *Chem. Educ. Res. Pract.* **2019**, DOI: 10.1039/C9RP00060G.

(34) Yan, F.; Talanquer, V. Students' Ideas about How and Why Chemical Reactions Happen: Mapping the Conceptual Landscape. *Int. J. Sci. Educ.* **2015**, *37* (18), 3066−3092.

(35) Weinrich, M.; Talanquer, V. Mapping Students' Conceptual Modes When Thinking about Chemical Reactions Used to Make a Desired Product. *Chem. Educ. Res. Pract.* **2015**, *16*, 561−577.

(36) Talanquer, V. Concept Inventories: Predicting the Wrong Answer May Boost Performance. *J. Chem. Educ.* **2017**, *94* (12), 1805−1810.

(37) Maeyer, J.; Talanquer, V. Making Predictions about Chemical Reactivity: Assumptions and Heuristics. *J. Res. Sci. Teach.* **2013**, *50* (6), 748−767.

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

(38) Talanquer, V. Explanations and Teleology in Chemistry Education. *Int. J. Sci. Educ.* **2007**, *29* (7), 853−870.
(39) Trommler, F.; Gresch, H.; Hammann, M. Students' Reasons for Preferring Teleological Explanations. *Int. J. Sci. Educ.* **2018**, *40* (2), 159−187.
(40) Abrams, E.; Southerland, S. The How's and Why's of Biological Change: How Learners Neglect Physical Mechanisms in Their Search for Meaning. *Int. J. Sci. Educ.* **2001**, *23* (12), 1271−1281.
(41) Weinrich, M. L.; Sevian, H. Capturing Students' Abstraction While Solving Organic Reaction Mechanism Problems across a Semester. *Chem. Educ. Res. Pract.* **2017**, *18* (1), 169−190.
(42) Sevian, H.; Bernholt, S.; Szteinberg, G. A.; Auguste, S.; Pérez, L. C. Use of Representation Mapping to Capture Abstraction in Problem Solving in Different Courses in Chemistry. *Chem. Educ. Res. Pract.* **2015**, *16* (3), 429−446.
(43) Kraft, A.; Strickland, A. M.; Bhattacharyya, G. Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11*, 281−292.
(44) Windschitl, M.; Thompson, J.; Braaten, M. Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Sci. Educ.* **2008**, *92* (5), 941−967.
(45) Bernholt, S.; Parchmann, I. Assessing the Complexity of Students' Knowledge in Chemistry. *Chem. Educ. Res. Pract.* **2011**, *12* (2), 167−173.
(46) Talanquer, V. Progressions in Reasoning about Structure−Property Relationships. *Chem. Educ. Res. Pract.* **2018**, *19* (4), 998−1009.
(47) van Mil, M. H. W.; Boerwinkel, D. J.; Waarlo, A. J. Modelling Molecular Mechanisms: A Framework of Scientific Reasoning to Construct Molecular-Level Explanations for Cellular Behaviour. *Sci. Educ.* **2013**, *22* (1), 93−118.
(48) Luisi, P. L. Emergence in Chemistry: Chemistry as the Embodiment of Emergence. *Found. Chem.* **2002**, *4* (3), 183−200.
(49) Cruz-Ramirez de Arellano, D.; Towns, M. H. Students' Understanding of Alkyl Halide Reactions in Undergraduate Organic Chemistry. *Chem. Educ. Res. Pract.* **2014**, *15*, 501−515.
(50) Flynn, A. B. Structure And Evaluation Of Flipped Chemistry Courses: Organic & Spectroscopy, Large And Small, First To Third Year, English And French. *Chem. Educ. Res. Pract.* **2015**, *16*, 198−211.
(51) Flynn, A. B. Flipped Chemistry Courses: Structure, Aligning Learning Outcomes, and Evaluation. In *Online Approaches to Chemical Education*; American Chemical Society: Washington, DC, 2017; pp 151−164.
(52) Flynn, A. B.; Ogilvie, W. W. W. Mechanisms before Reactions: A Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow. *J. Chem. Educ.* **2015**, *92* (5), 803−810.
(53) Ogilvie, W. W.; Ackroyd, N.; Browning, S.; Deslongchamps, G.; Lee, F.; Sauer, E. *Mechanistic Patterns*, 1st ed.; Nelson Education Ltd.: Toronto, ON, 2017.
(54) Deslongchamps, G. *Organic Chemistry Flashware*; Nelson Education Ltd., 2007.
(55) Seufert, T. Supporting Coherence Formation in Learning from Multiple Representations. *Learn. Instr.* **2003**, *13* (2), 227−237.
(56) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books: Sausalito, CA, 2006.
(57) IUPAC. Hammond Principle (Hammond Postulate). *IUPAC Gold Book*; 2006; https://goldbook.iupac.org/html/H/H02734.html (accessed Apr 2019).
(58) Krippendorff, K. Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educ. Psychol. Meas.* **1970**, *30* (1), 61−70.
(59) Hallgren, K. A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor. Quant. Methods Psychol.* **2012**, *8* (1), 23−34.
(60) Krippendorff, K. Computing Krippendorff's Alpha-Reliability. http://repository.upenn.edu/asc_papers/43 (accessed Apr 2019).
(61) DeFever, R. S.; Bruce, H.; Bhattacharyya, G. Mental Rolodexing: Senior Chemistry Majors' Understanding of Chemical and Physical Properties. *J. Chem. Educ.* **2015**, *92* (3), 415−426.
(62) Brown, N. J. S.; Nagashima, S. O.; Fu, A.; Timms, M.; Wilson, M. A Framework for Analyzing Scientific Reasoning in Assessments. *Educ. Assess.* **2010**, *15* (3−4), 142−174.
(63) Johnson-Laird, P. N. Mental Models and Human Reasoning. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (43), 18243−18250.
(64) Popova, M.; Bretz, S. L. Organic Chemistry Students' Interpretations of the Surface Features of Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 919−931.
(65) Popova, M.; Bretz, S. L. Organic Chemistry Students' Challenges with Coherence Formation between Reactions and Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 732−745.
(66) Sandoval, W. A. Conceptual and Epistemic Aspects of Students' Scientific Explanations. *Journal of the Learning Sciences.* **2003**, *12* (1), 5−51.
(67) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.* **2018**, *95* (2), 207−217.
(68) Talanquer, V. Exploring Dominant Types of Explanations Built by General Chemistry Students. *Int. J. Sci. Educ.* **2010**, *32* (18), 2393−2412.
(69) Pazicni, S.; Bauer, C. F. Characterizing Illusions of Competence in Introductory Chemistry Students. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 24−34.
(70) Berland, L. K.; Reiser, B. J. Making Sense of Argumentation and Explanation. *Sci. Educ.* **2009**, *93* (1), 26−55.

DOI: 10.1021/acs.jchemed.8b00719
J. Chem. Educ. 2019, 96, 1068−1082

Chapter 3: Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items

3.1 Introduction to the research article presented in this chapter

In the research described in Chapter 2, we learned how students constructed scientific arguments related to the judging the plausibility of an organic chemistry reaction. This analysis revealed to us how students constructed arguments in one organic chemistry context (*i.e.*, mechanisms), but did not tell us much about how students' argumentation patterns might be similar or different in different contexts. Research is needed to understand how learners are engaging in argumentation in different contexts so that we can determine what components of different activities might shape how students to reason with evidence. We could then use these findings to guide construction of assessments and activities to support students' argumentation skills. Chapter 3 describes our study on how students' argumentation patterns may vary in different and change between content areas and tasks. Using the same reasoning, granularity, and comparisons framework from Chapter 2, we analyzed and compared arguments from students to two organic chemistry prompts from a single final exam: (a) deciding and justifying which of two organic chemistry mechanisms was more plausible (n = 159); and (b) deciding and justifying which of three bases could drive an acid–base equilibrium to products (n = 170).

This study built on the research in Chapter 2, which was led by Nick Bodé and Dr. Flynn. This was my first experience carrying out a study from start to finish, and I definitely struggled at this stage to get a sense of the related literature, research questions/methods, and scope/goals of the study. As a junior graduate student who was getting adjusted to a new degree, institution, and home, everything was overwhelming. I think what got me over the hump was the collective guidance and support from Dr. Flynn and my colleagues (particularly: Dr. Amanda Bongers, Dr. Myriam McKenna, and Keith Lapierre). Through the many analyses, revisions, and instances of

sheer demoralization, they always reminded me of the question that guides all the Flynn group's work: "Why is this important?". I think it was this framing that helped me zoom out and think about the big picture, which ultimately helped us get this project to the finish line. I want to thank Dr. Flynn for her support and patience, as well as Dr. Myriam McKenna for her help with the inter-rater reliability.

3.2 Copyright

Reproduced from "Deng, J. M.; Flynn, A. B. Reasoning, Granularity, and Comparisons in Students' Arguments on Two Organic Chemistry Items. *Chem. Educ. Res. Pract.* **2021**, *22*, 749–771. doi.org/10.1039/D0RP00320D" with permission from the Royal Society of Chemistry.

Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Educ. Res. Pract.*, 2021, **22**, 749

Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items

Jacky M. Deng and Alison B. Flynn *

In a world facing complex global challenges, citizens around the world need to be able to engage in scientific reasoning and argumentation supported by evidence. Chemistry educators can support students in developing these skills by providing opportunities to justify how and why phenomena occur, including on assessments. However, little is known about how students' arguments vary in different content areas and how their arguments might change between tasks. In this work, we investigated the reasoning, granularity, and comparisons demonstrated in students' arguments in organic chemistry exam questions. The first question asked them to decide and justify which of three bases could drive an acid–base equilibrium to products (Q1, n = 170). The majority of arguments exhibited relational reasoning, relied on phenomenological concepts, and explicitly compared between possible claims. We then compared the arguments from Q1 with arguments from a second question on the same final exam: deciding and justifying which of two reaction mechanisms was more plausible (Q2, n = 159). The arguments in the two questions differed in terms of their reasoning, granularity, and comparisons. We discuss how course expectations related to the two questions may have contributed to these differences, as well as how educators might use these findings to further support students' argumentation skill development in their courses.

Received 20th October 2020,
Accepted 5th April 2021

DOI: 10.1039/d0rp00320d

rsc.li/cerp

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Introduction

Citizens need to be able to argue from scientific evidence

In a world facing complex global issues (United Nations, 2015), citizens need to be able to make decisions and argue for those decisions using scientific evidence. For example, an evidence-based decision of whether to vaccinate requires deciding to rely on evidence (rather than intuition and emotion), interpreting the quality of the available evidence, and using this evidence to reason for or against a particular decision (Jones and Crow, 2017).

National frameworks for science education in the United States have identified explanations and arguments about phenomena as a key scientific practice (National Research Council, 2012), and the importance of such skills has also been articulated in Europe (European Union, 2006; Jimenez-Aleixandre and Federico-Agraso, 2009), Canada (Social Sciences and Humanities Research Council, 2018), and other international organizations (*e.g.*, Organisation for Economic Cooperation and Development, 2006). However, chemistry education research has found that opportunities for students to argue and explain have largely been absent within traditional chemistry assessments. For example, constructing scientific explanations appeared in less than 10% of American Chemical Society (ACS) general chemistry exam items examined in 2016 (Laverty *et al.*, 2016; Reed *et al.*, 2017). Additionally, an ACS Exam for organic chemistry did not assess students' ability to construct scientific explanations or arguments at all (Stowe and Cooper, 2017). To better support student development of argumentation and explanation skills, curricula have emerged that explicitly include argumentation and explanation (Talanquer and Pollard, 2010; Cooper and Klymkowsky, 2013), as well as research focused on characterizing argumentation and explanation in laboratory settings (Carmel *et al.*, 2019).

Arguments provide insight into students' reasoning

Arguments and explanations are distinct. An explanation is used to explain an agreed-upon fact or phenomenon (Osborne and Patterson, 2011; National Research Council, 2012), while arguments justify a fact or phenomenon that is not agreed-upon (McNeill *et al.*, 2006; Kuhn, 2011); rather, the claim is in doubt and must be advanced through reasoning by constructing an argument about the fit between the evidence and claim (Toulmin, 1958; Osborne and Patterson, 2011). Arguments therefore provide a an opportunity to investigate how students are reasoning about phenomena (Emig, 1977; Berland and Reiser, 2009; Grimberg and Hand, 2009).

Department of Chemistry & Biomolecular Sciences, University of Ottawa, 10 Marie Curie, Ottawa, Ontario, K1N 6N5, Canada. E-mail: alison.flynn@uOttawa.ca

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Recent studies in chemistry education research have worked to characterize students' reasoning by analysing their arguments about chemical phenomena (Sevian and Talanquer, 2014; Weinrich and Talanquer, 2016; Bodé *et al.*, 2019; Moon *et al.*, 2019; Moreira *et al.*, 2019). For example, Sevian and Talanquer (2014) interviewed individuals ranging from high school chemistry students to chemistry experts (*e.g.*, academia, and industry professionals). The interviewees were asked to construct arguments when deciding on a fuel to power a GoKart; through their responses, the researchers characterized students' reasoning as one of descriptive, relational, linear causal, or multi-component causal. These modes of reasoning have since been used in other studies to characterize students' reasoning through analysis of arguments and explanations across a variety of contexts and tasks (Sevian and Talanquer, 2014; Weinrich and Talanquer, 2016; Bodé *et al.*, 2019; Moon *et al.*, 2019; Moreira *et al.*, 2019). In the present study, we analyzed students' reasoning in an acid–base context.

Acid–base equilibria are key to many domains of chemistry

Knowledge of acid–base chemistry underpins understanding of the majority of reactions in both organic chemistry and biochemistry, and previous work found that acid–base reactions are often the first reaction type taught to organic chemistry students (Stoyanovich *et al.*, 2015).

Research on acid–base chemistry concepts has identified how many students struggle with the Brønsted–Lowry and Lewis definitions of acids and bases, applying Lewis acid–base chemistry in novel contexts (Bhattacharyya, 2006; Cartrette and Mayo, 2011; McClary and Talanquer, 2011), describing why acid–base reactions proceed in the fashion that they do (Cooper *et al.*, 2016), and interpreting and using data related to acid–base chemistry, such as pK_a and pH data (Krajcik and Nakhleh, 1994; Orgill and Sutherland, 2008; Flynn and Amellal, 2016).

Previous research has also sought to identify students' misconceptions about individual chemical equilibrium concepts, such as Le Chatelier's principle and chemical equilibrium equations (Wheeler and Kass, 1978; Hackling and Garnett, 1985; Banerjee, 1991; Quilez-Pardo and Solaz-Portoles, 1995; Huddle and Pillay, 1996; Voska and Heikkinen, 2000). However, work is needed that directly investigates students' competencies using acid–base concepts within the context of chemical equilibria, as the synthesis of these two domains of chemistry underpins many of the phenomena students encounter in biochemical and biological contexts later in their studies (*e.g.*, enzymes, ocean acidification).

Given the foundational role that acid–base chemistry plays in other reactivity, we first sought to investigate how students construct an argument within the context of an acid–base equilibrium. This content area has also yet to be investigated within the current chemistry education literature on argumentation, despite its relative importance in both general and organic chemistry and biochemistry (Duis, 2011).

Studies focused on argumentation in chemistry education have also been limited to single content areas (*i.e.*, investigating students' arguments for a single question), which has made it difficult to determine how different tasks might influence students' arguments. For example, students may struggle to generate sophisticated arguments in one content area but may not struggle in other content areas. Therefore, in this work, we next compared students' arguments in the acid–base question with a previous analysis of students' arguments in a different content area: comparing mechanistic pathways (Bodé *et al.*, 2019).

Analytical framework

We analysed students' arguments using a framework with three dimensions: modes of reasoning, granularity, and comparisons (Fig. 1), as described below.

Modes of reasoning. Reasoning has been analysed through a variety of different lenses and frameworks in chemistry education research. These approaches include Type I and II reasoning (Talanquer, 2007, 2017; McClary and Talanquer, 2011; Maeyer and Talanquer, 2013), teleological reasoning (Talanquer, 2007;

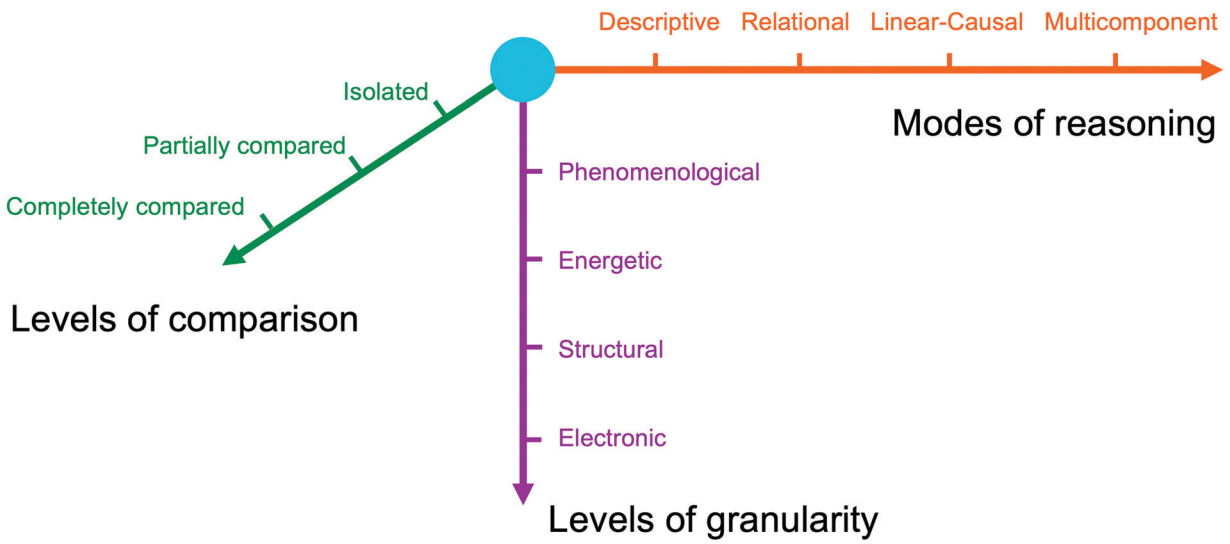


Fig. 1 The dimensions comprising the analytical framework in this work: reasoning, granularity, and comparisons.

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Abrams and Southerland, 2010; Trommler *et al.*, 2018; Caspari *et al.*, 2018a, 2018b; DeCocq and Bhattacharyya, 2019), abstractedness and abstraction (Sevian *et al.*, 2015; Weinrich and Sevian, 2017), rules-, case-, and model-based reasoning (Windschitl *et al.*, 2008; Kraft *et al.*, 2010; DeCocq and Bhattacharyya, 2019), and causal, mechanistic, and causal mechanistic reasoning (Cooper *et al.*, 2016; Crandell *et al.*, 2018).

In this study, we analysed students' arguments in terms of four modes of reasoning: descriptive, relational, linear causal, and multi-component causal (Sevian and Talanquer, 2014; Weinrich and Talanquer, 2016; Caspari *et al.*, 2018a). We chose this framework because of its alignment with the intended learning outcomes of the course context in which this study was conducted, including the associated classroom activities related to crafting scientific arguments. We describe each mode below.

Descriptive arguments list or give features and/or the properties of entities (*e.g.*, the reactants, products) without establishing connections. For example, to justify a claim that humans are causing global warming, one might give "Burning fossil fuels generating CO_2." However, without an explicit link between the evidence and the claim, it is unclear how the evidence is connected to the claim, if at all (*e.g.*, Why is CO_2 important in climate change? How does it have an effect?).

Relational arguments include connections between properties of the entities and their activities, but these relationships are discussed in a correlative fashion (*i.e.*, absent of causality). In other words, connections are stated but the argument does not extend to *why* these links or evidence are appropriate. For example, to justify a claim that humans are causing global warming, one might state: "Humans are causing global warming because they generate CO_2 by burning fossil fuels." Compared to the descriptive example, this argument includes an *explicit* link between the evidence and the claim. However, the reader is left wondering why or how CO_2 contributes to global warming.

Causal arguments include all features of a relational argument and additionally contain cause-and-effect relationships between the relevant properties of the entities and their activities. In other words, links are stated and additional reasoning explains why or how these links are relevant and/or appropriate, often by referencing scientific knowledge, principles, additional evidence, *etc.* *Linear causal* arguments establish a single chain of causal relationships between one or more pieces of evidence to justify a claim. For example, a linear causal argument to justify a claim that humans are causing global warming may be: "Humans are causing global warming because they generate CO_2 by burning fossil fuels. CO_2 is a greenhouse gas that contributes to global warming by trapping heat in the Earth's atmosphere." Here, the second sentence serves as the reasoning that explains the relationship between the claim and evidence in the first sentence.

Multi-component causal arguments establish multiple chains of causal relationships between more than one piece of evidence to support a claim. A multi-component causal argument to justify the claim that humans are causing global warming may include the same linear causal example above, but with an added "chain" of causal reasoning to support the original claim, such as: "Humans are causing global warming because they generate CO_2 from burning fossil fuels. CO_2 is a greenhouse gas that contributes to global warming by trapping heat in the Earth's atmosphere. In addition, humans participate in agricultural activities that increase CH_4 generation, another greenhouse gas that traps heat in the Earth's atmosphere." The argument could continue even further by describing the chemical properties of CO_2 and CH_4 that make them greenhouse gases, a concept that we describe below as levels of granularity.

Levels of granularity. Beyond constructing arguments about phenomena at different levels of reasoning, they can be constructed at different levels of granularity (Fig. 2). For example, a justification for why aspirin acts as an acid in water may focus on pH and pK_a data (a phenomenological level) or how aspirin has a carboxylic acid functional group that is resonance-stabilized when deprotonated (underlying level that includes structural, electronic, and energetic factors) (Talanquer, 2018a). Different contexts and tasks require different levels of granularity, as different phenomena may be explained from increasingly large macroscopic perspectives (*e.g.*, global levels and beyond) or increasingly small submicroscopic perspectives (*e.g.*, atomic levels and beyond) (Darden, 2002). The idea of granularity has been described in other work on scientific reasoning, including scales (Talanquer, 2018b), levels (van Mil *et al.*, 2013), nested hierarchies (Southard *et al.*, 2017), emergence (with ideas of downward and upward causality) (Luisi, 2002), and bottom-out reasoning (Darden, 2002).

In this study, we categorized students' arguments into four levels of granularity relevant to the questions they were asked: phenomenological, energetic, structural, and electronic.

The phenomenological level captures descriptions of chemical phenomena that arise from the interactions of molecules and atoms and their structural, electronic, and energetic properties. For example, within a given context, the favoured direction of a chemical equilibrium may be a phenomenon to be explained. The interplay of structural, electronic, and energetic properties/interactions of molecules and atoms can be used to determine and justify the direction of an equilibrium. Depending on the task, arguments may also be focused on other phenomenological factors that can be equally valid; for example, pK_a data could be used to determine the direction of an acid–base equilibrium.

The structural level captures descriptions of structural features of molecules and atoms. For example, in an acid–base equilibrium context, a student's description of the relative stability of two basic atoms would be considered discussion at the structural level. In the context of an organic chemistry mechanism, a structural discussion might include identifying steric bulk around a particular reactive centre and connecting the steric interactions to the effects at the transition state (energetic level). The structural level itself contains grain sizes, such as cells, biomolecules, small molecules, molecular fragments, functional groups, and individual atoms.

The electronic level captures descriptions of electronic features of molecules and atoms. For example, electronegativity and partial charges could be used to explain reactivity at the electronic level. Other examples might compare formal charges and electron density on basic atoms to justify the direction of

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

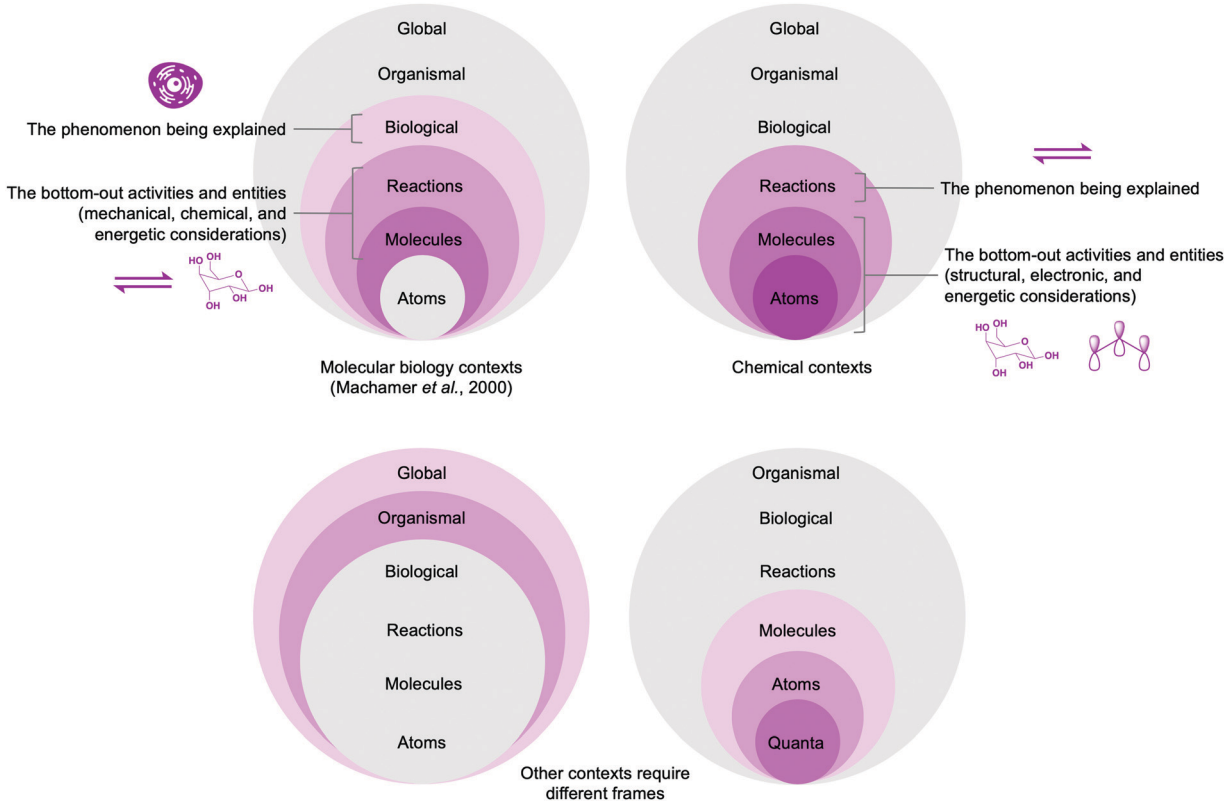


Fig. 2 Different contexts/tasks require different levels of granularity.

an equilibrium, or discuss molecular orbitals to describe electronic features of molecules.

The energetic level captures descriptions of the energetics of reactions, including thermodynamic and kinetic considerations. Descriptions at this level could include considering the relative stabilities of conjugate acids/bases to justify the direction of an equilibrium or justifying the plausibility of various reaction mechanisms based on activation energies.

In this study, we used these four levels of granularity based on the concepts and ideas identified in students' responses, the intended learning outcomes related to the questions we analysed, as well as previous theoretical work related to chemistry students' reasoning (Machamer *et al.*, 2000; Luisi, 2002; van Mil *et al.*, 2013; Southard *et al.*, 2017; Talanquer, 2018a). Different levels of granularity may be more relevant for other contexts, such as other content areas within chemistry or other disciplines (biology, physics, *etc.*). For example, chemical reactions and equilibria may be the phenomena to be explained in the chemical contexts investigated in this study and the highest level of granularity needed for these contexts, while in molecular biology contexts, these phenomena may be the deepest level of granularity needed for an explanation (*e.g.*, explaining why a substrate binds to an enzyme).

Levels of comparison. A comparison is needed when an argument involves two or more possible claims, or when there are various factors that influence an outcome, phenomenon, or claim (Toulmin, 1958). Without a comparison, a species cannot be more/less, bigger/smaller, or faster/slower than another. Comparing between alternatives is also a key aspect of scientific practice; for example, to justify why global warming is happening, one might leverage evidence to refute counterclaims (claims that global warming does not exist). In the questions used in this study, students had to argue for one of multiple claims, thereby providing an opportunity to construct arguments in which they compared their claim to alternatives. The arguments may include full, partial, or no comparisons.

Goals and research questions

We characterized students' arguments for an acid–base equilibrium question (Q1) in terms of the concepts, links, and comparisons that were articulated, specifically using the following research question (RQ):

(1) When constructing an argument to decide which base will drive an equilibrium towards products:

(a) What concepts do students include?

(b) What links do students establish between concepts?

(c) What concepts do students use to compare between claims?

(d) What modes of reasoning do the arguments exhibit?

Next, we used the findings from RQ1 to compare with the analysis of a question that prompted students to compare mechanisms (Q2) (Bodé *et al.*, 2019). Specifically, we investigated:

(2) How might students' arguments differ on two different organic chemistry questions from a single exam in terms of reasoning, granularity, and comparisons?

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Methods

Setting and course

This research was conducted within the context of an Organic Chemistry II course at a large, bilingual, research-intensive university in Canada. At this institution, introductory organic chemistry is offered across two semesters as Organic Chemistry I (OCI) and Organic Chemistry II (OCII). OCI is offered in the winter semester of students' first year of studies while OCII is offered in both the following summer and fall. Students can take the courses in either English or French. OCII is a 12 week course (~400 students per section) consisting of two weekly classes (1.5 hours each, mandatory, lecture or flipped format) and a voluntary tutorial session (1.5 hours) (Flynn, 2015, 2017). Assessments for the course are comprised of in-class participation *via* a classroom response system, online homework assignments, two midterms, and a final exam. The course is comprised of ~75% Faculty of Science students, ~17% Faculty of Health Sciences students, and ~8% students from other faculties. General topics addressed in OCII include reactions with σ electrophiles (*i.e.*, S_N1/S_N2/E1/E2 and oxidation reactions), introduction to ^{1}H NMR and IR spectroscopy, reactions of π electrophiles with leaving groups, and reactions with activated π nucleophiles (*e.g.*, aldol reactions) (Flynn and Ogilvie, 2015; Ogilvie *et al.*, 2017; Raycroft and Flynn, 2020).

Data source. We analysed and compared findings from students' responses to two final exam questions (Fig. 3) from the OCII 2017 final exam. Question 1 (Q1, n = 170) asked students to justify the direction of an acid–base equilibrium and Question 2 (Q2, n = 159) asked students to justify why one of two similar reaction mechanisms was more plausible (S_N1 *versus* S_N2). For Q1, pK_a values were not provided to students, though values for chemical analogues were provided in a data table attached to the exam. Each question followed Toulmin's claim–evidence–reasoning pattern, as students were asked to: (a) choose a claim given multiple options, (b) justify their choice in an argument using evidence and reasoning. Prior to our analysis, we received Research Ethics Board approval (H03-15-18).

The analysis of concepts, links, comparisons, and modes of reasoning for Q2 had been previously reported as part of a separate research work (Bodé *et al.*, 2019). We used this analysis to support our investigation of RQ2. Therefore, when discussing concepts, links, and comparisons in this work (RQ1), we report only the analysis and findings for Q1.

Coding process. The first part of our analysis focused on the concepts, links, and comparisons in students' arguments. We initially identified these components based on the expected answer to Q1 (Appendix B), which was constructed based on the intended acid–base learning outcomes from the OCII course (Appendix C). This process established content validity for the initial coding scheme, ensuring that we defined the initial scheme using concepts that matched course expectations. During the coding process, we added codes that were not present in the initial coding scheme but were present in students' answers. We included these additional codes even if they were described in error or were irrelevant to the question.

The analysis followed the following sequence:

(1) Identifying concepts present in the argument and whether these concepts were discussed correctly or with errors.

(2) Identifying links between individual concepts in the argument and whether these links were canonically correct or not.

(3) Identifying which concepts were used to explicitly compare/contrast between possible claims.

Only explicit instances of concepts were coded. For example, we only coded for the concept of "base strength" if the argument included phrases like "NaH is a strong base". Links between concepts were said to be present only when the student was explicitly linking between concepts with words like "because", "therefore", "so", *etc.* A concept was said to compare between claims if reference was made to one or more of the other possible claims. For example, "NaH is a stronger base than NH_3" or "NaH is a strong base and NH_3 is a weak base" would warrant a comparison code for a *base strength* concept code. Next, we determined the mode of reasoning of students' arguments to be one of *descriptive*, *relational*, *linear causal*, or *multi-component causal* using the definitions provided in Table 1. For example, a descriptive argument was defined as one in which a student simply described concepts or features of molecules but did not make any connections between these statements (*e.g.*, stating a claim and providing some evidence, but not connecting these ideas). In contrast, linear-causal response was said to be present if a student made a claim (*e.g.*, "The equilibrium will favour products…"), justified that claim with a concept/feature ("because NaH is a strong base…"), and justified this connection by describing how we know *why* a strong base drives the equilibrium towards products ("A strong base drives the equilibrium towards products because it has a conjugate acid with the highest pK_a value"). Appendix A provides additional examples of the coding process for the modes of reasoning.

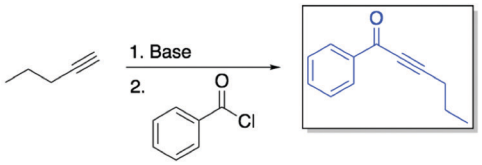
Canonical correctness of the links was not a factor when deciding on the mode of reasoning, as (a) we were principally interested in students' domain-general abilities to reason and (b) an argument can still be logically sound while including canonically incorrect information (Toulmin, 1958).

To support our analysis, we drew diagrams to visually represent students' arguments. These diagrams allowed us to visually organize the units (links and concepts) within students' arguments, helping us assign a mode of reasoning to each argument. Examples of diagrams to facilitate analysis of arguments have been previously described (Verheij, 2003; Moreira *et al.*, 2019) and we provide examples of diagrams used in this work in Appendix A.

We assigned a level of granularity to each argument based on the granularity of the concepts identified in the first part of the analysis (Table 2). For example, in Q1, we categorized an argument relating two concepts—direction of an equilibrium and pK_a values—to be at the phenomenological level of granularity because this argument did not consider any underlying factors that contributed to these phenomena (*i.e.*, it did not acknowledge any energetic, structural, or electronic factors). In contrast, an answer that discussed how the electronegativity of a particular

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Question 1 (Q1): Acid–Base Equilibrium



a. Draw the major product of the reaction in the box above.

b. Circle the base below than can be used to force the equilibrium of the first step to the product side:

Claim

NaOH NH$_3$ **NaH**

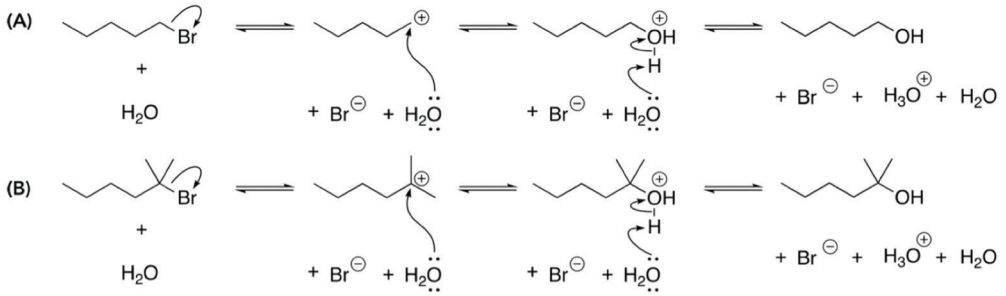
c. Explain your answer in part b (why you chose one base **and** did not choose the others), using chemical structures as part of your answer.

Evidence

Reasoning

Question 2 (Q2): Comparing Mechanisms

a. Draw and label the reaction coordinate diagram for the two reactions below that are shown proceeding by an S$_N$1 mechanism.



b. Which reaction is most likely to proceed by the mechanism shown? **B**

Claim

c. Justify your answer in part b, using your reaction coordinate diagram as part of your explanation.

Evidence

Reasoning

Fig. 3 The acid–base equilibrium question (Q1, top), and the comparing mechanisms questions (Q2, bottom). Both questions prompted students for their claim, evidence, and reasoning.

atom (electronic) could be used to determine the relative stability of the molecule (energetic) and the direction of an equilibrium (phenomenological) was coded as having discussed concepts at three distinct levels of granularity.

Lastly, we coded each argument as one of three levels of comparison—*isolated*, *partially compared*, and *fully compared*—based on the degree to which concepts in the argument were used to compare between the possible claims (Table 3). For example, if an argument included the concepts *base strength* and *acid strength*, but both these codes were discussed only in terms of the chosen claim, then we coded this argument as *isolated*. If one (but not both) of these concepts was used to compare to another possible claim ("NaH is a stronger base than NH_3"), then we coded this argument as *partially compared*. If both concepts were used to compare to another base ("NaH is a stronger base than NH_3, which means H_2 is a weaker

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Table 1 Descriptions and examples of the modes of reasoning

Mode of reasoning	Description	Example
Descriptive	Claims are supported solely by describing features and/or the properties of entities (*e.g.*, the reactants, products, *etc.*). No connections are established	*NaH is the strongest base*
Relational	Relationships between properties of the reaction materials and their activities are discussed in a correlative fashion (*i.e.*, a cause–effect relationship is not established)	*NaH will drive the equilibrium towards products because its conjugate acid has a higher pK_a value (35) than the alkyne (25)*
Linear causal	All of the features of a relational argument are present and are accompanied by cause-and-effect relationships between the relevant properties of the entities and their activities	*NaH will drive the equilibrium towards products because its conjugate acid (pK_a value 35) is weaker/more stable than the starting alkyne (pK_a value 25), according to the relevant pK_a values. NaOH and NH_3 are not strong enough bases to drive the equilibrium to the product side, as both of their conjugate acids are stronger than the starting alkyne (pK_a values of 15.7 and 9, respectively)*
Multi-component causal	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct inter-actions of several components	*The hydride is the only base of the choices that is sufficiently strong to drive the equilibrium to products. The hydride base is smaller and less electronegative than carbon, making it less able to stabilize a negative charge than the alkyne anion (conjugate base)* *The oxygen atom in NaOH is more electronegative than the carbon atom in the conjugate base, making it better able to stabilize the negative charge. Although atomic size (larger carbon atom) and hybridization suggest that the carbon atom in the conjugate base could be more stable, the pK_a values of the acid and conjugate acids support the conclusion that NaOH is the weaker base, since the stronger the acid, the weaker its conjugate base (pK_a value of H_2O = 15.7; pK_a of alkyne = 25)* *NH_3 is similarly too weak to drive the equilibrium to the product side, with the nitrogen atom being neutral in addition to being more electronegative than the carbon anion*

Table 2 Examples of concepts at each level of granularity for Q1 and Q2. Concepts with a * indicate ones that students proposed in their responses but were unexpected based on course learning outcomes

Examples of concepts at each level of granularity		
Level of granularity	Q1: acid–base equilibrium	Q2: comparing mechanisms
Phenomenological	Direction of equilibrium, pK_a values, conjugate acid strength, base strength, *reaction pathway	*Reaction likelihood
Structural	Atom size	Steric hinderance, number of α-carbon substituents
Electronic	Electronegativity, formal charge, *nucleophilicity	Hyperconjugation
Energetic	Relative stability	Activation energy

Table 3 Descriptions for each level of comparison from Bodé *et al.* (2019)

Level of comparison	Description
Isolated	Concepts in argument for a claim are all discussed in isolation from the other possible claim. Concepts are never used to compare/contrast between the claims
Partially compared	Some (but not all) concepts in argument for a claim are discussed in relation to the other possible claim. These concepts are used to compare/contrast between the claims
Fully compared	All concepts in argument for a claim are discussed in relation to the other possible claim. All concepts are used to compare/contrast between the claims

conjugate acid than NH_4^+''), then we coded this statement as *fully compared*.

Inter-rater reliability

To evaluate and improve the reliability of our qualitative analysis, a second coder analysed a subset of exams for the units outlined in Table 4 using the method described above to establish inter-rater reliability (Krippendorff, 1970; Hallgren, 2012). We used Krippendorff's α as a statistical measure to evaluate agreement between coders (Krippendorff, 1970). Unlike percent agreement, Krippendorff's α accounts for chance agreement between coders. We calculated inter-rater reliability for the analysis of concepts, links, comparisons, and modes of reasoning, as levels of granularity and levels of

Table 4 Krippendorff α values obtained from inter-rater analysis for units in students' arguments. Acceptable agreement = 0.67

	Krippendorff's α		
Unit of analysis	Round 1	Round 2	Round 3
Concepts	0.58	0.77	0.82
Links between concepts	0.42	0.71	0.86
Concepts used to compare	0.58	0.86	0.95
Mode of reasoning	0.40	0.49	0.74

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

comparison were dependent on concepts and comparisons, respectively.

For each question, after the primary coder coded the entire set of responses, the second coder used the first iteration of the codebook to code a subset of 15% of students' arguments. Both coders then met to discuss differences between their respective analyses. The most common challenges in our coding were (1) determining whether a student was making implied references to links or comparisons and (2) determining the arguments' mode of reasoning. For example, one argument stated "NaH is the strong base. The equilibrium is forced to the products." In this case and similar cases, the coders were unsure about the presence/absence of implied links and comparisons. Based on these discussions, we decided to code mainly for explicit references to links and comparisons to limit the number of assumptions we could make during our analysis. Any assumptions about implied references were first discussed with other raters before making a final decision. We repeated the interrater process with new subsets of data (15% of the dataset) until the two coders obtained a Krippendorff's α greater than 0.67 for each of the units described in Table 4, the value that exceeds the threshold of acceptability for inter-rater reliability (Krippendorff, 1970). Between each round of the inter-rater process, the codebook (Appendix A) was revised based on discussions between the two raters.

Results and discussion

The following sections related to RQ1 describe findings from our analysis of the concepts, links, and comparisons identified in students' arguments to Q1. We had collected similar data for Q2 in previous work (Bodé *et al.*, 2019) and used this previously collected data for investigating RQ2.

RQ1a: What concepts do students include?

For Q1, we found differences in the concepts discussed depending on whether they provided a correct or incorrect claim (Fig. 4). Arguments with correct claims more frequently discussed the *direction of the equilibrium*, *conjugate acid strength*, and the *pK_a values of conjugate acids*. In the context of the OCII course, all three of these concepts were relevant to the claim and were key concepts employed in the expected answer for this question (Appendix B).

For incorrect claims, the two most frequently discussed concepts were *base strength* and *reaction pathways*. For example, Student 10 provided the following argument which used base strength to justify a suggested reaction pathway:

Student 10: "NaH is **the strong base** choice therefore it is most likely to react by **deprotonating the carbon**." [emphasis by the authors]

Despite *base strength* being the most prevalent concept discussed in incorrect claims, the majority of arguments for incorrect claims discussed this concept incorrectly. This was found to be reflective of a broader trend, as correct claims were more frequently justified with concepts that were discussed correctly compared to incorrect claims.

RQ1b: What links do students establish between concepts?

We visualized the links made between concepts in students' arguments for Q1 using Gephi data visualization software. Nodes represent concepts; edges (*i.e.*, a line between two nodes) represent links between two concepts (Fig. 5). The frequency of links between two concepts is correlated with thickness of the edge. In other words, a thicker edge represents two nodes (concepts) that were more frequently connected in students' arguments. In contrast, a node with no edges represents a concept that had no links to other concepts in the dataset.

Three concepts were the most prevalent in correct claims: the *direction of the equilibrium*, *conjugate acid strength*, and *pK_a values of conjugate acids*. These were also the three concepts that exhibited the most frequent connections. Often, arguments for correct claims included a triad of concepts and links that included stating the respective pK_a values of the conjugate acids of the given bases, using these pK_a values to rank the relative strengths of the conjugate acids, then using these rankings to justify the extent to which an equilibrium involving each base/conjugate acid would favour a particular direction. For example, Student 116 provided the following argument which included this triad:

Student 116: "I chose NaH as a base because its conjugate acid has a **pK_a value of around 36**, which makes it **a weaker acid**

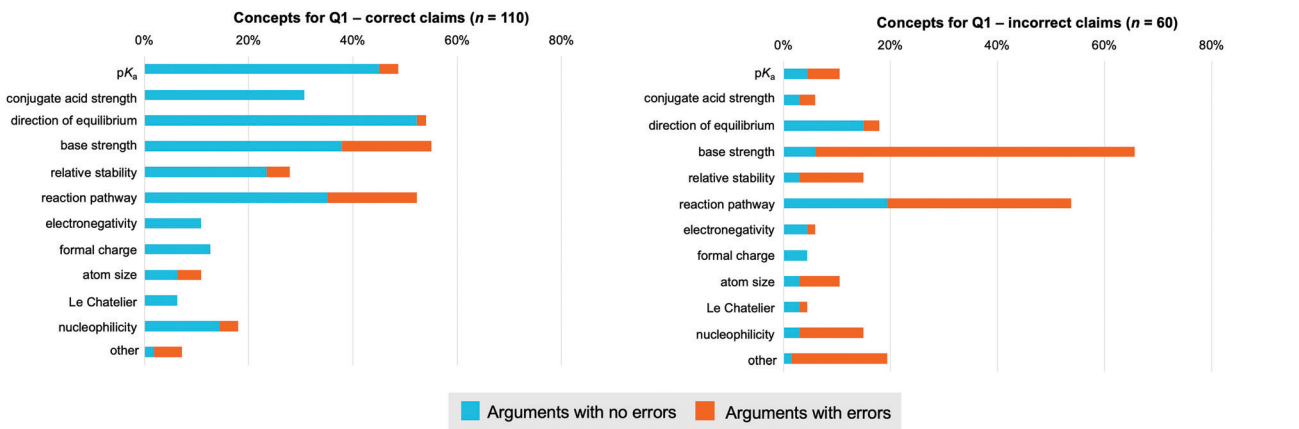
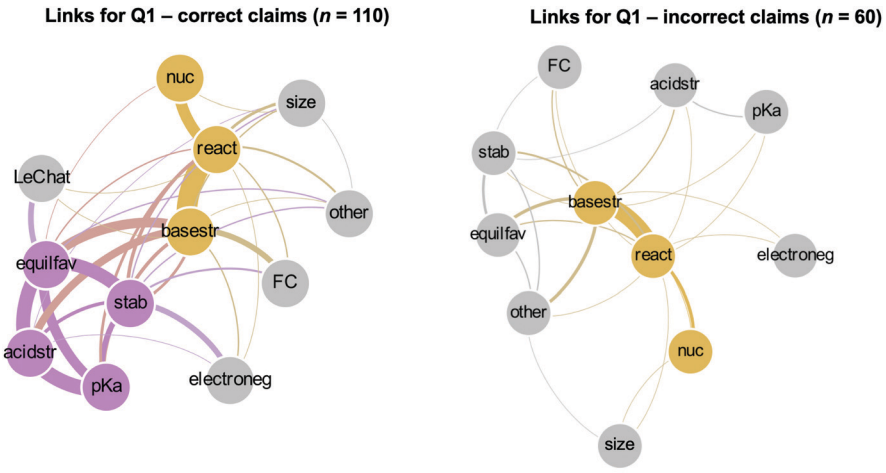


Fig. 4 For Q1, concepts discussed in arguments for correct claims (n = 110, left) and incorrect claims (n = 60, right).

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.



other	Other	**react**	Reaction pathway
nuc	Nucleophilicity	**stab**	Stability
LeChat	Le Chatelier	**basestr**	Base strength
size	Atom size	**equilfav**	Direction of the equilibrium
FC	Formal charge	**acidstr**	Conjugate acid strength
electroneg	Electronegativity	**pKa**	pK_a value of conjugate acid

Fig. 5 For Q1, connections made between concepts made for correct claims (left, n = 110) and incorrect claims (right, n = 60).

than the starting material. The **equilibrium will favour the side with the weaker acid**. I did not choose NaOH or NH_3 because their respective conjugate acids would have a **pK_a value less than that of the SM [starting material], meaning that the equilibria would favour the starting materials (pK_a ~ 15.7 for H_2O and ~10 for NH_4^+).**''

In some cases, this type of argument was expanded to include a discussion of *base strength*. These arguments included identifying the relationship between the relative strengths of the conjugate acids from the relative strengths of the bases, and then using these ideas in concert to determine the direction of the equilibrium.

The most common connection made in incorrect claims was between *base strength* and *reaction pathway*. In these cases, students often used base strength as the principle concept to justify how their chosen base (or all three bases) would react with the alkyne or the acyl chloride. For example, Student 43 provided the following argument, which linked NaOH being a strong base to how the base would proceed in a reaction (compared to the other options):

Student 43: ''[NaOH is] a **strong base** that can **remove the hydrogen from the alkyl chain**, whereas the other bases are **weaker** and need more activation energy **to remove the hydrogen**.''

We suspect that students who linked the codes *base strength* and *reaction pathway* may have done so in a rote fashion. This link was present in both incorrect claims and correct claims; however, in correct claims, *base strength* was also linked to other concepts, such as *conjugate acid strength*.

RQ1c: What concepts do students use to compare between claims?

Fig. 6 shows how often a given concept was used in a comparison between claims. Correct claims primarily compared between claims during discussions of *pK_a values of conjugate acids*, *conjugate acid strength*, and *the direction of the equilibrium*. For example, Student 14 listed the pK_a values for all three conjugate acids, used these to compare the relative strength of the acids based on these values, then described which direction the equilibrium would favour in each case:

Student 14: ''I chose NaH as the base because **its conjugate acid has a higher pK_a value** than the alkyne. That means that the conjugate acid is a weak acid, weaker than the alkyne, so the reaction will favour the products. I did not choose NaOH or NH_3 because **their conjugate acids have smaller values than the alkyne**, driving the equilibrium towards the starting materials.''

In contrast, incorrect claims primarily compared between claims using the concepts of *base strength* and *reaction pathways*. A common example was a student stating that one base was stronger than the other two bases, leading them to conclude that the stronger base would be able to react as a base with the alkyne. For example, Student 55's argument:

Student 55: ''NaH will take the H of the bonding end of the triple bond to make H_2(g). **NaH is a much stronger base than NaOH and NH_3.** NaOH and NH_3 are **too weak to deprotonate the alkyne. NH_3 would break the triple bond and add NH_2 to the end of the triple bond. NaOH wouldn't react at all.** NaH when a solution has H^- floating around, which are extremely reactive.''

Determining the levels of comparison for Q1, arguments for correct claims more frequently compared against the other

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

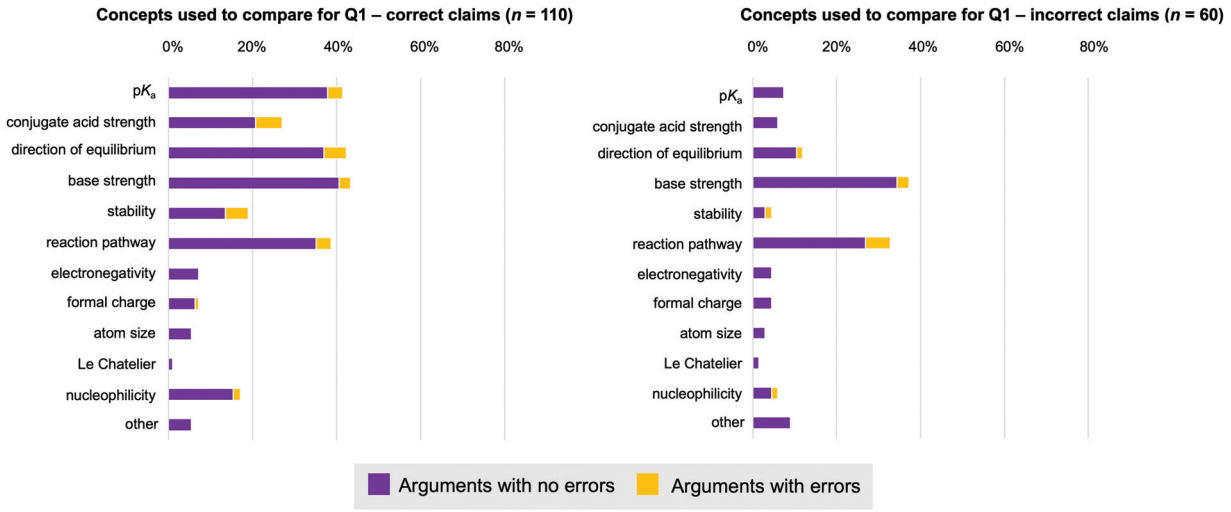


Fig. 6 For Q1, how often each concept was used to compare between claims in arguments for both correct (left, n = 110) and incorrect (right, n = 60) claims.

possible claims than arguments for incorrect claims, $\chi^2(1, N = 170) = 11.2$, $p = 0.001$, $\phi = 0.257$ (Fig. 7). In other words, students who provided correct claims were more likely to compare and contrast between claims, while students who provided incorrect claims were more likely to discuss their claim in isolation of the other possible claims.

RQ1d: What modes of reasoning do the arguments exhibit?

For Q1, the majority of students (62%) provided the correct claim (*i.e.*, chose the correct base) for which base would drive the equilibrium in question to products (Fig. 8). However, causal reasoning was present in only 31% of all answers (either linear causal or multi-component causal). Correct claims more frequently exhibited causal arguments than incorrect claims (linear causal and multi-component causal), while incorrect claims more frequently exhibited descriptive arguments than correct claims. The frequency of causal arguments was significantly different between arguments for correct claims *vs.* arguments for incorrect claims, $\chi^2(1, N = 170) = 18.1$, $p < 0.001$ with a medium effect size, $\phi = 0.33$.

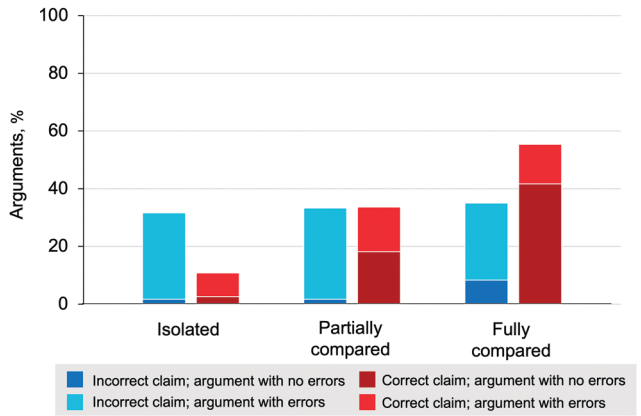


Fig. 7 Levels of comparison for Q1. Students who provided correct claims (n = 110) were more likely to compare and contrast between claims, while students who provided incorrect claims (n = 60) were more likely to discuss their claim in isolation of the other possible claims.

Relational arguments were the most prevalent across all student arguments for Q1 (48% of all answers). The most common relational argument discussed how a chosen base was a strong base (*base strength*) that was strong enough to drive the equilibrium towards products (*direction of the equilibrium*). Other relational arguments were similar but discussed acid strength or pK_a values in place of base strength. The commonality here was that these arguments did not include discussions of *why* base strength, acid strength, or pK_a values would affect the direction of the equilibrium. In contrast, a common linear causal argument discussed how the equilibrium would favour the products due to differences in pK_a values and would then explain why these pK_a values were relevant to the claim by referencing how pK_a values enabled comparison between relative acids strengths. For example, the first part of Student 19's argument linked the direction of the equilibrium to conjugate acid strength, and justified this link with pK_a values:

Student 19: "The equilibrium of the first step is dependent on the acid–base reaction and as a result, it is **dependent on which side does** [sic] **the stronger acid lie.** Based on the structure of the reactant, **the more acidic proton is at the terminal alkyne (pK_a 50 [C–H sp^3] *vs.* 24 [C–H sp]), so the appropriate base must have a weaker conjugate acid**..."

Although this argument is linear causal, it has a phenomenological level of granularity, as there is no discussion of any underlying factors that contribute to acid strength/pK_a values and the direction of an equilibrium. The latter portion of Student 19's argument does achieve a deeper level of granularity by relating these phenomena to electronic factors, such as electronegativity:

Student 19 (continued): "...Based on the electronegativity of OH and NH_3, they would serve as better bases than the alkyne as **the greater electronegativity of O and N allowing the ionized forms to better stabilize a negative charge** (for O, making the $^-$OH a more stable base than the ionized alkyne) and less able

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

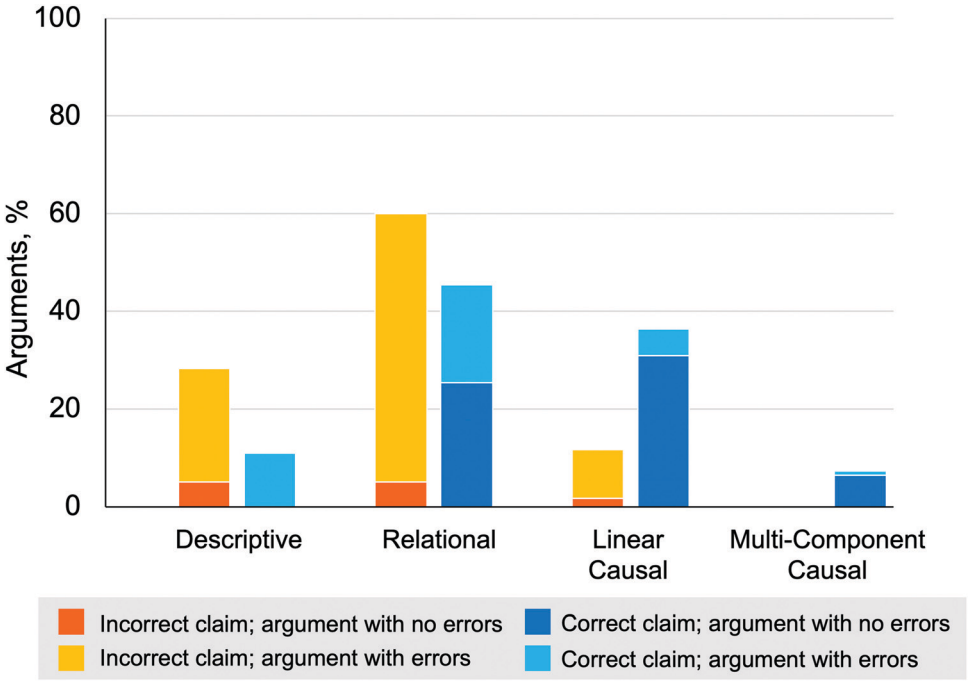


Fig. 8 Modes of reasoning for students' arguments in Q1 (correct claims, n = 110; incorrect claims, n = 60). Students who were arguing for correct claims were more likely to exhibit causal modes of reasoning.

to stabilize a positive charge (for N, NH_4^+ (CA for NH_3) is more acidic than alkynes and hence, shifts equilibrium to the alkyne)."

Multi-component causal arguments were only present in arguments for correct claims. The most common multi-component causal arguments justified the direction of the equilibrium using both base concepts (*base strength*, *electronegativity*) and acid concepts (*conjugate acid strength*, pK_a *values*).

RQ2: How might students' arguments differ on two different organic chemistry questions from a single final exam in terms of reasoning, granularity, and comparisons?

The distributions for the modes of reasoning for Q1 arguments differ qualitatively from the modes of reasoning for Q2 arguments uncovered in our previous work (Bodé *et al.*, 2019) (Fig. 9). To determine the statistical significance of these differences, we compared the respective percentages of causal and non-causal arguments between Q1 and Q2 to determine the extent to which students' reasoning differed between the two questions. We found that arguments for Q2 had significantly more causal arguments than for Q1 (linear and multi-component), with a medium effect size, $\chi^2(1, N = 329) = 20.456$, $p < 0.001$, $\phi = 0.27$.

Next, we determined the levels of granularity using the concepts identified in arguments for both Q1 and Q2. Each level of granularity had a different number of underlying concepts *(e.g.*, for Q1, five concepts were considered phenomenological,

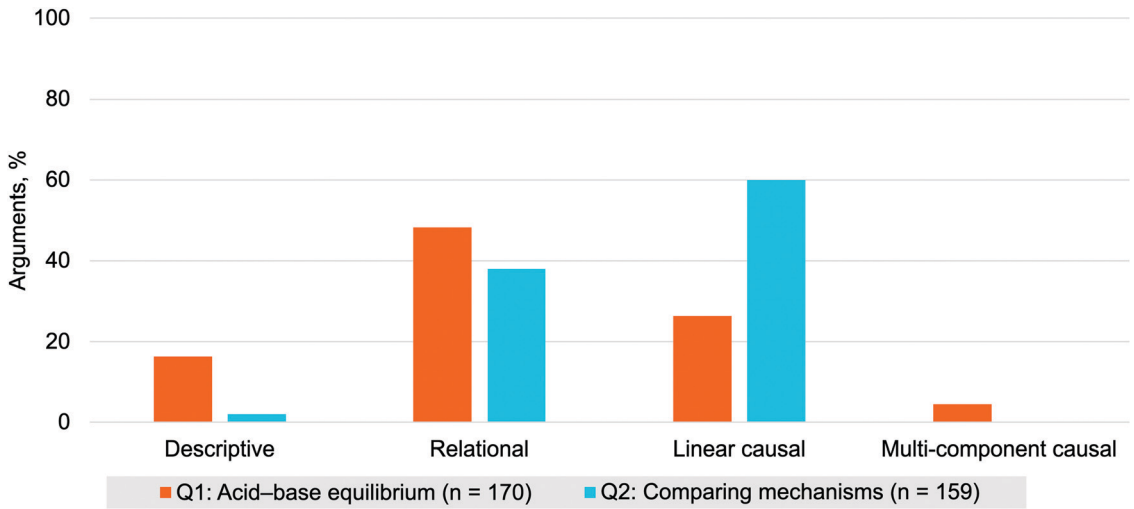


Fig. 9 Modes of reasoning for the acid–base equilibrium (Q1, n = 170) and comparing mechanisms (Q2, n = 159) questions.

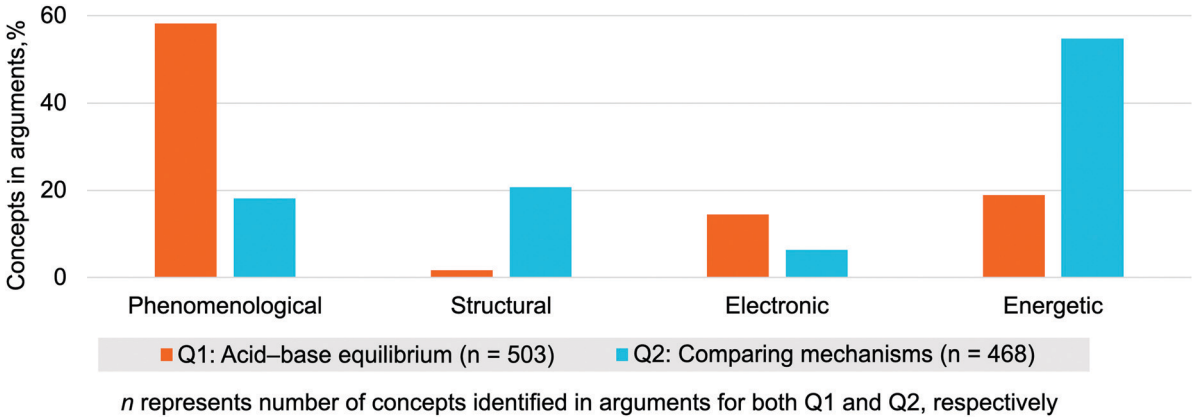


Fig. 10 The proportion of concepts exhibited at each level of granularity for both Q1 (acid–base, n = 503) and Q2 (comparing mechanisms, n = 468). Descriptions for each level of granularity are described in Table 2.

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

while only two concepts were considered electronic). We therefore normalized the different number of concepts that could be described at each level of granularity by dividing the frequency of concepts at each level of granularity by the number of possibilities for each level (*e.g.*, for Q1: dividing the sum of all concepts at phenomenological level by five).

Because Q1 and Q2 assessed different conceptual knowledge and required different levels of granularity, we qualitatively compared the granularity expressed in students' arguments for the two questions (Fig. 10). For Q1, the concepts were primarily at a phenomenological level of granularity (*e.g.*, arguments focused on *pK_a values, conjugate acid strength, direction of the equilibrium*); however, some students' arguments included concepts from more granular levels (*e.g.*, *electronegativity*, *formal charge*, *stability*). For Q2, the majority of concepts were at the structural and energetic levels, which included concepts such as number of *α-carbon substituents*, *number of carbocation substituents*, and *activation energy*.

We also investigated how students compared between claims in Q1 *versus* Q2 (Fig. 11). Students more frequently compared concepts (either partially or fully) on Q2 than Q1, $\chi^2(1, N = 329) = 10.748$, $p = 0.001$, $\phi = 0.18$. Additionally, when investigating the relative frequencies of partial *versus* full comparisons, we found that students more frequently made full comparisons on Q2 than Q1, $\chi^2(1, N = 329) = 36.170$, $p < 0.001$, $\phi = 0.354$.

We were interested in identifying potential factors for why a single group of students produced arguments that differed in terms of reasoning, granularity, and comparisons on a single exam. Therefore, we compared the intended and enacted learning outcomes from the OCII course for the two questions (Stoyanovich *et al.*, 2015; Raycroft and Flynn, 2020). Intended learning outcomes (ILOs) are defined as the knowledge, skills, and values students are *expected* to demonstrate by the end of a course (Biggs and Tang, 2011), which are often described in course syllabi. We analysed ways in which the ILOs were taught, practiced, and assessed through the course (Dixson and Worrell, 2016; Carle and Flynn, 2020; Raycroft and Flynn, 2020). First, we reviewed the OCII course syllabus for ILOs relevant to Q1 and Q2 (full list available Appendix C). We then reviewed how these ILOs were enacted in the course notes and videos (taught), problem sets and in-class activities (practiced), and midterms and exams (assessed).

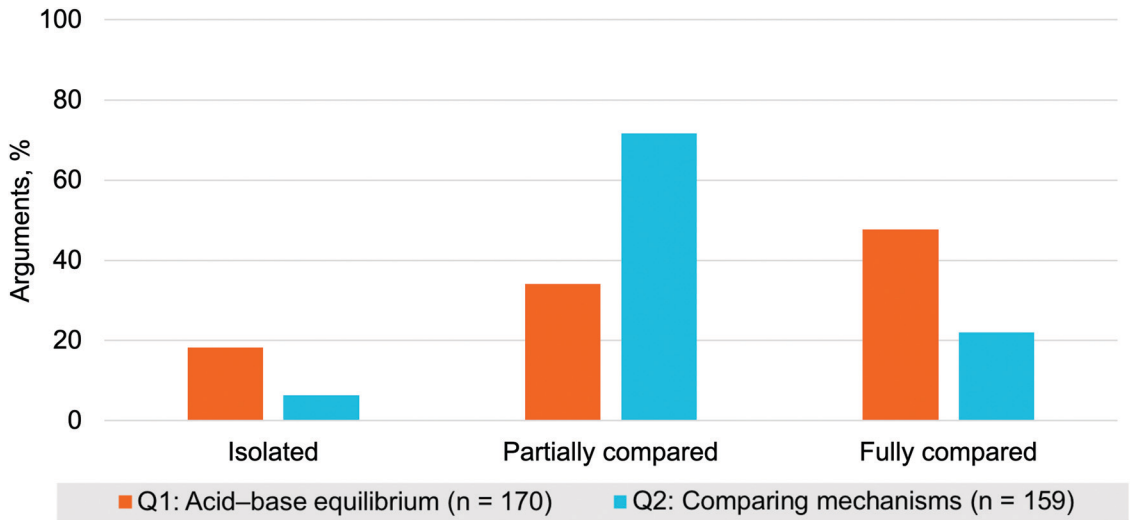


Fig. 11 Differences in the levels of comparison between Q1 (acid–base equilibrium, *n* = 170) and Q2 (comparing mechanisms, *n* = 159).

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Reviewing the course materials related to Q1 and Q2, we found that how these questions were taught, practiced, and assessed aligned well with how students responded to these questions. For Q1, students were expected throughout the course to be able to justify the direction of acid–base equilibria using *both* chemical factors and pK_a data (Flynn; Stoyanovich *et al.*, 2015; Flynn and Amellal, 2016). However, in cases where chemical factors were competing—for example, a base in the starting materials being resonance stabilized but the conjugate base in the products bearing a larger and more electronegative atom—students could rely on pK_a data of the acids (*i.e.*, experimental evidence) to make their final decision. This is the case in Q1, as orbitals/hybridization suggests that NaH is more stable than the acetylene anion, but electronegativity and charge suggest the opposite. Therefore, students may have focused their arguments on pK_a data to come to a final decision, perhaps resulting in the less granular, non-causal arguments found in our analysis.

In contrast, for Q2, students were expected to leverage a combination of structural and energetic information when making decisions about whether S_N1/S_N2 and E1/E2 reactions would occur (examples of course notes in Appendix C). Further, on the midterm exam earlier in the course, students had been asked a question similar to Q2 of this study in which they were expected to justify which of two mechanisms was more plausible by establishing connections between the structural features of molecules and energetic information within reaction coordinate diagrams. These activities may have reinforced expectations throughout the class about generating more granular, causal arguments for questions like Q2, such as those we uncovered in our analysis.

Conclusions

This study provides insight into how students construct arguments when justifying the direction of an acid–base equilibrium (RQ1) as well as how students' arguments can differ between content areas within chemistry (RQ2). This work adds to a growing body of research on analysing students' abilities to justify claims about chemical phenomena through argumentation and reasoning.

For Q1, arguments for correct and incorrect claims were focused on different sets of concepts; arguments with correct claims more frequently discussed pK_a values and conjugate acid strength while arguments with incorrect claims more frequently discussed relative base strength and described how molecules would react (RQ1a). Arguments for correct claims more frequently linked the direction of the equilibrium to pK_a values, conjugate acid strength, and relative base strength, while incorrect claims more frequently linked relative base strength to descriptions of how molecules would react (RQ1b). Arguments for correct claims more frequently completely compared between different bases in their arguments, while incorrect claims more frequently discussed claims in isolation of other possibilities. Lastly, arguments with correct claims more frequently exhibited causal reasoning (linear causal and multicomponent causal), while incorrect claims more often exhibited relational reasoning.

Related to the second research question (RQ2), Q1 arguments demonstrated more relational reasoning compared to Q2 arguments, which demonstrated more causal reasoning. In general, concepts discussed in Q1 were more phenomenological, often focusing on pK_a values or general descriptors (*e.g.*, strong acid, strong base) to justify claims. In comparison, arguments for Q2 more often argued using underlying factors, such as structural and energetic information, to justify their claims. Lastly, Q1 arguments exhibited more complete comparisons between claims than Q2 arguments.

Students' arguments on the two questions broadly aligned with how these questions were taught, practiced, and assessed within the course context (Fig. 12). These findings reinforce the notion that students' arguments—including the reasoning, granularity, and comparisons demonstrated in an argument, as shown in this work—depend on the course context, the stakes, how well expectations are communicated, in addition to students' actual abilities (Kelly *et al.*, 1998; Sadler, 2004; Sadler and Zeidler, 2005; von Aufschnaiter *et al.*, 2008; Barwell, 2018; Cian, 2020). For example, research on students' arguments in other content areas in organic chemistry, such as delocalization, have also found that students' arguments can differ depending on the task/context (Carle *et al.*, 2020). In summary, from Q1 and Q2 combined, over 60% of students in this work demonstrated that they *can* construct causal arguments, but whether they choose to will depend on appropriateness and need (Bodé *et al.*, 2019).

Implications for teaching and research

If we expect students to argue in a particular way and leverage specific concepts and/or evidence in their arguments, then as educators we need to be explicit and consistent in how we communicate these expectations within our course contexts (Fig. 12) (Bernholt and Parchmann, 2011; Stoyanovich *et al.*, 2015; Weinrich and Talanquer, 2016; Caspari *et al.*, 2018b; Carle and Flynn, 2020). As noted by Macrie-Shuck and Talanquer (2020): "the complex nature of mechanistic reasoning in chemistry demands integrating multiple pieces of knowledge and connecting various scales (*e.g.*, macro, multiparticle, single-particle), dimensions (compositional, energetic), and modes of description and explanation (phenomenological, mechanical, structural). Developing mastery in this area likely demands time and sustained and concerted effort across multiple courses and areas of knowledge." Although causal arguments are suggested to be more sophisticated modes of reasoning in various frameworks used to characterize reasoning, this mode of reasoning is not necessarily "better" than any other mode. What is considered "better" reasoning depends on the argument's context and purpose; in scientific practice and chemical thinking, less "sophisticated" arguments may be completely acceptable, practical, and successful for accomplishing a given task and meeting a certain expectation.

One potential avenue to further investigate the influence of course context and expectations on students' arguments might be asking students to construct two arguments, each with a different mode of reasoning, level of granularity, and level of

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

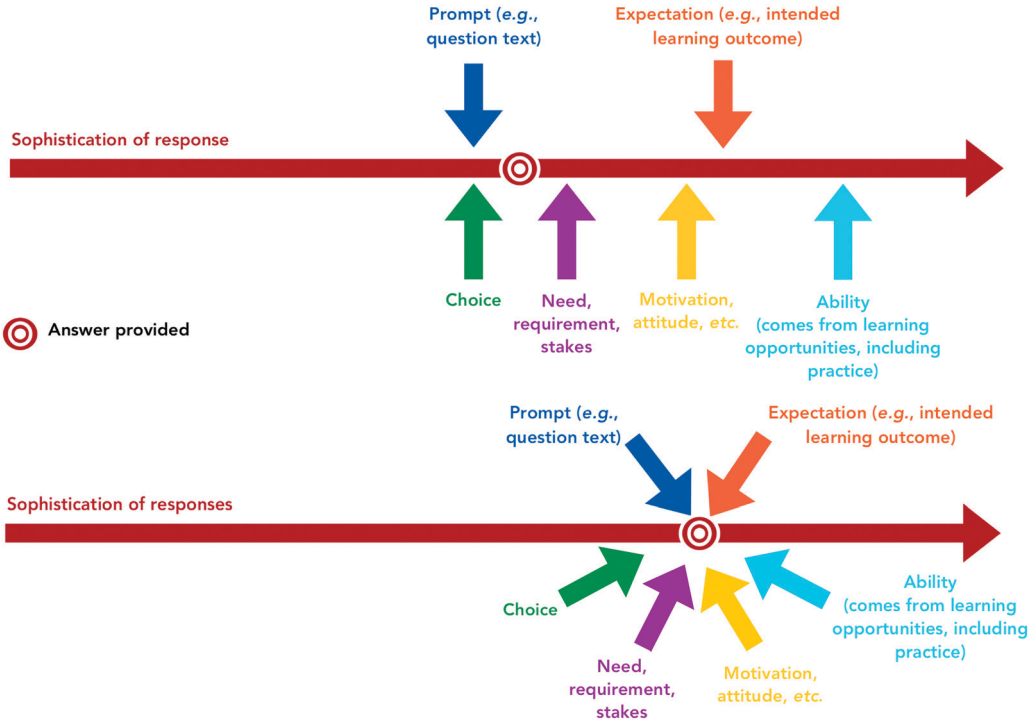


Fig. 12 Aligning different factors within a course context can help support student achievement of the intended learning outcomes.

comparison, respectively, and to determine whether students are able to effectively traverse across these dimensions when constructing arguments. Another option would be to provide students with pre-constructed arguments and ask them to identify the reasoning, granularity, and comparison(s). For example, the OCII course has incorporated assessment items that explicitly prompt students to consider the different chemical factors and pK_a data involved in making decisions about chemical equilibria (Fig. 13).

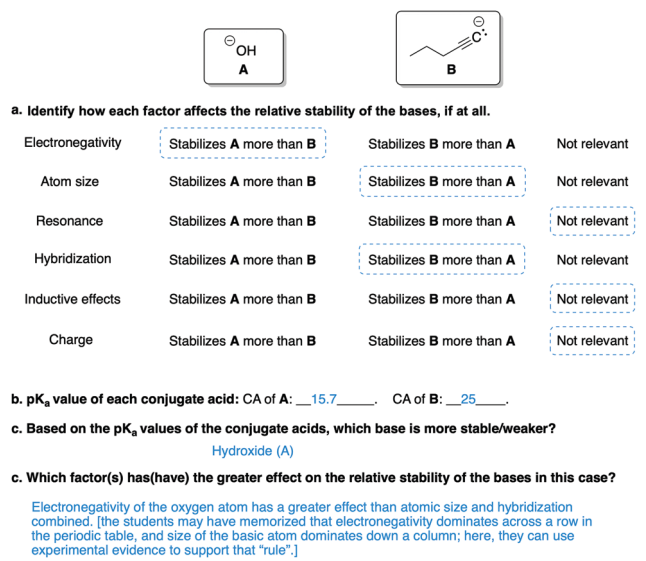


Fig. 13 Example of assessment item to prompt students to consider factors at various levels of granularity.

Limitations

Open responses such as the ones analysed in this study provide rich insight into students' thinking; however, they are limited in that they may give an incomplete picture. For example, the design of the prompts presented in this work may have influenced the types of responses students generated (*e.g.*, no multicomponent reasoning exhibited in Q2). We decided to analyse students' written arguments in this work to allow for statistical analysis of trends within a larger sample. Other qualitative methods such as interviews and focus groups would provide researchers with even richer insight and more opportunities for dialogue and inquiry.

Conflicts of interest

There are no conflicts to declare.

Appendix A: codebook and coding examples

List of concepts (Table 5)

Concepts are defined as key words or phrases in a student's argument that indicate their use of a chemical property, concept, phenomenon, *etc.*

Coding sequence for concepts, links, and comparisons

1. Concepts

For each concept, the following considerations should be made:

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Table 5 Codes for concepts proposed in arguments, with notes to guide appropriate application

Code	Concept	Notes
pK_a	pK_a value of conjugate acids	Reference to the pK_a value of any (or all) of the conjugate acids described • May include explicit discussion of relative pK_a values without referring to actual values (*e.g.*, pK_a value of H_2O is lower than that of H_2) • Correct if the pK_a value for the chosen base is stated incorrectly or discussed incorrectly relative to other bases • Incorrect if the pK_a for the chosen base is stated incorrectly or discussed incorrectly relative to other bases
Acidstr	Conjugate acid strength	Reference to the strength of the conjugate acid(s) of any of the bases provided • Must explicitly discuss acid strength of the conjugate acid(s) (using terms like "strong" and/or "weak") • Do not code if argument only mentions "conjugate acid" without reference to the strength of that conjugate acid
Equilfav	Direction of equilibrium	Reference to the direction favoured in an acid–base equilibrium • Correct if the favoured direction is aligned with other concepts (stability, pK_a, *etc.*) • Incorrect if the favoured direction is misaligned with other concepts (stability, pK_a, *etc.*)
Basestr	Base strength	Reference to the strengths ("strong", "weak") of any of the three bases • Correct if NaH > NaOH > NH_3 discussed explicitly or implicitly ○ Explicit: NaH is a stronger base than NaOH or NH_3 ○ Implicit: NaH is the strongest base • Incorrect if comparisons between bases are incorrect
Stab	Stability of molecule	Reference to the stability of a molecule • Does not need to include detailed discussion of energetics; simply saying "stable" warrants inclusion of this code • Also code for references to favourability of lower-energy species ("H_2 is a low energy species") • Correct if the conclusion about stability is aligned with other concepts (acid strength, pK_a, *etc.*) • Incorrect if the conclusion about stability is misaligned with other concepts (acid strength, pK_a, *etc.*)
React	Reaction pathway	Describing how a molecule will react • Example: "Because NaH is a strong base, **it will be able to deprotonate the alkyne**" • Correct if the correct reactive pathway is described (base deprotonating acid) • Incorrect if incorrect reactive pathway is described (nucleophilic attack of carbonyl, *etc.*)
Electroneg	Electronegativity	Reference electronegativity of the basic atoms on any of the given bases • Code if differences in electron density are discussed or if an atom is said to have a greater capacity to attract electrons; argument does not need to explicitly say "electronegativity" • Correct if O > N > H discussed explicitly or implicitly
FC	Formal charge	Reference to the charges (or lack of charges) on any of the bases • Must be explicitly discussed as a concept in-text ("The oxygen has a negative charge, which means…") • Correct if appropriate charge assigned to species
Size	Atom size	Reference to the atomic size of any of the bases • Correct if B > A > C discussed explicitly or implicitly • B > A, C acceptable. For example, "NH_3 is the largest base of the three" would be coded as correct • Incorrect if comparisons between bases are incorrect
LeChat	Le Châtelier's principle	References to the equilibrium "shifting" to compensate for loss of gaseous products Correct if phrases describing Le Châtelier's principle is used. No need to say "Le Châtelier's principle" explicitly in the argument
Other Nuc	Other Nucleophilicity	Concepts that are unique and infrequent. Provide supplementary notes to define what "other" is referring to Discussing the nucleophilicity of bases • Can be discussed correctly even though nucleophilicity not relevant to the argument • Correct if A, B > C discussed explicitly or implicitly (C is less nucleophilic and A, B are more nucleophilic) • Incorrect if comparisons between bases are incorrect

• Is it discussed correctly [y(g)] (y = yes, concept is present; g = correct) or with errors [y(e)] (y = yes, concept is present; e = error)?

• Note that there may be cases where the discussion of the concept itself is correct while the link to another concept is incorrect

○ Example: *NaH is the strongest base because it is the smallest of the three bases.*

– Relative base strength and atom size are each discussed correctly.

2. Links

• Is the basis for the link correct (g) or incorrect (e)?

• In most cases, a link will be established in an argument as a student describes how Concept 1 impacts or explains Concept 2.

• If two concepts are linked, but the argument moved through *an additional concept* to link them, be sure to code all three concepts as being linked.

○ Example: if someone states that pK_a is linked to conjugate acid strength which is then linked to base strength, do not code pK_a as being linked to base strength (unless the two ideas are explicitly linked elsewhere). Links in this case would be p*K*a – acidstr – basestr.

3. Comparisons

• Example: "NaH is the strongest base" or "NaH is a stronger base than NaOH which is a stronger base than NH_3"

• If no comparisons are made, still include the base for which the concept is being discussed in reference to.

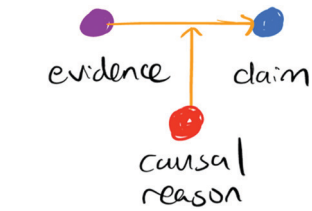
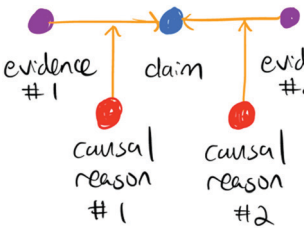
○ If the argument only discusses the pK_a of A, then the comparison portion of the code would simply have "A"

In a given cell in an Excel spreadsheet (in this example, coding for an instance of the concept "base strength"):

basestr
y(g/e) – [linked concept 1](g/e),…,[linked concept *n*](g/e) – NaH to NaOH

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

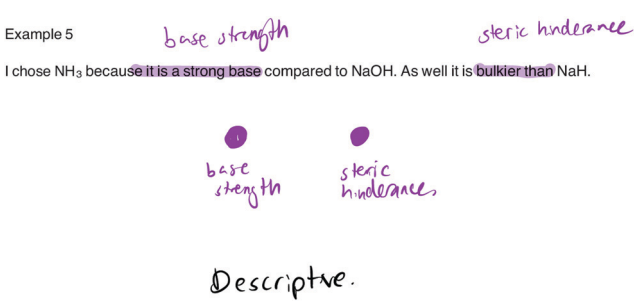
Table 6 Descriptions of the four modes of reasoning and associated diagrams

Level of reasoning	Description	Example diagram
Descriptive (D)	Claims are supported solely by describing features and/or the properties of entities (*e.g.* the reactants, products, *etc.*)	
Relational (R)	Relationships between properties of the reaction materials and their activities are discussed in a correlative fashion (i.e., a cause–effect relationship is not established)	
Linear causal (L)	All of the features of a relational argument are present and are accompanied by cause-and-effect relationships between the relevant properties of the entities and their activities	
Multi-component causal (M)	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct interactions of several components	

Coding Examples

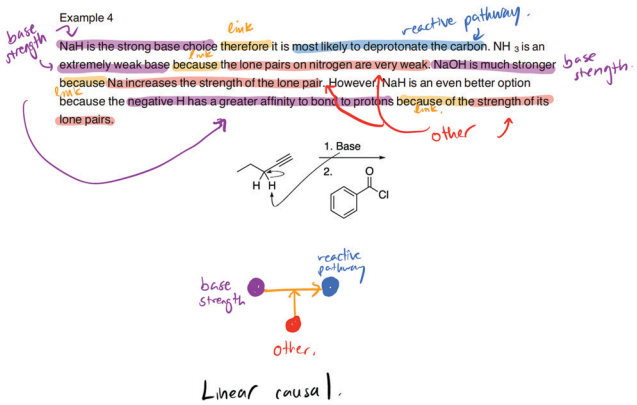
Example 1

Concepts: Base strength discussed incorrectly. Steric hinderance (coded as *other* in this context) discussed correctly.
Level of granularity: Base strength is a phenomenological concept. Steric hinderance is a structural concept.
Links: No links established between any concepts.
Mode of reasoning: Descriptive. No explicit links established between any of the concepts described.
Comparisons: NH_3 explicitly stated to be stronger base than NaOH and bulkier than NaH.
Level of comparison: Completely compared. Both concepts are explicitly used to compare between the bases.



Example 2

Concepts: Base strength discussed correctly. ''Strength of lone pairs'' considered as other and discussed incorrectly. A reactive pathway is described but it is incorrect.
Level of granularity: Base strength is a phenomenological concept. Describing a reactive pathway is phenomenological.
Links: Base strength is linked to the reactive pathway and this link is justified through ''strength of lone pairs'' (other).
Mode of reasoning: Linear causal. The suggested reactive pathway is said to be dependent on base strength, and this is justified with ''strength of lone pairs'' (other).
Comparisons: Base strength and ''strength of lone pair'' explicitly compared between all three bases. The reactive pathway suggested is only discussed in relation to NaH.
Level of comparison: Partially compared.



 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

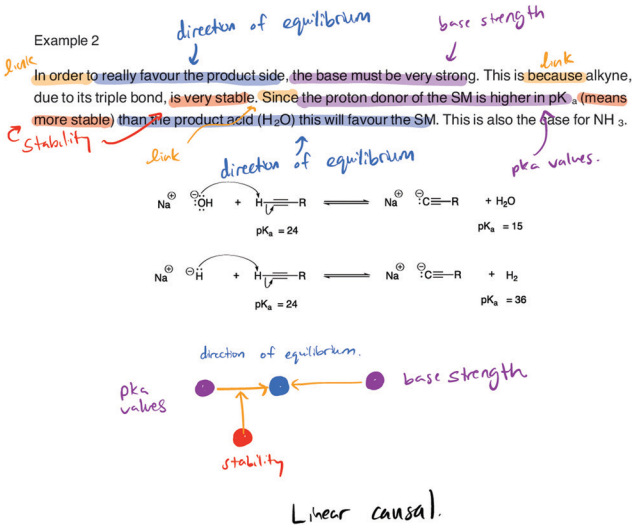
Example 3

Concepts: pK_a values, conjugate acid strength, base strength, direction of the equilibrium, electronegativity, stability, and Le Chatelier's principle all discussed correctly.
Level of granularity: pK_a values, conjugate acid strength, base strength, direction of the equilibrium, and Le Chatelier's principle are all phenomenological. Electronegativity is an electronic concept. Stability is an energetic concept.
Links: The direction of the equilibrium is linked to pK_a values, conjugate acid strength, base strength, electronegativity. Stability is linked to Le Chatelier's principle and electronegativity.
Mode of reasoning: Multi-component causal. The direction of equilibrium is justified in three distinct ways: (1) pK_a values, (2) conjugate acid strength, (3) electronegativity with reference to how this influences stability, and (4) Le Chatelier's principle with reference to how this influences stability.
Comparisons: pK_a values and conjugate acid strength are only discussed for NaH. Base strength and electronegativity are used to compare NaOH and NH_3. Only NaH is included in discussions about Le Chatelier's principle. All three bases are compared in discussions involving the direction of the equilibrium and stability.
Level of comparison: Partially compared.



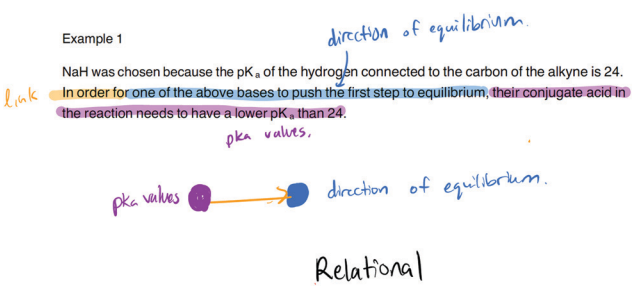
Example 4

Concepts: pK_a values, base strength, direction of the equilibrium, and stability all discussed correctly.
Level of granularity: pK_a values, base strength, and direction of the equilibrium, are all phenomenological. Stability is an energetic concept.
Links: The direction of the equilibrium is linked to pK_a values and base strength. Stability is linked to pK_a values.
Mode of reasoning: Linear causal. The direction of the equilibrium is said to be dependent on pK_a values, and this link is justified with reference to stability. Base strength is also offered as a factor, but no causal reasoning provided for why this is relevant to the direction of an equilibrium.
Comparisons: All concepts described are discussed in a general comparison across all three claims.
Level of comparison: Fully compared.



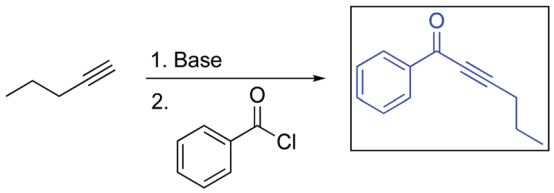
Example 5

Concepts: pK_a values and direction of the equilibrium both discussed correctly.
Level of granularity: pK_a values and direction of the equilibrium are both phenomenological.
Links: The direction of the equilibrium is linked to pK_a values.
Mode of reasoning: Relational. The direction of the equilibrium is said to be dependent on pK_a values, but this is link is not justified with additional causal reasoning.
Comparisons: Concepts are used to discuss NaH only.
Level of comparison: Isolated



Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Appendix B: expected answer based on intended acid–base learning outcomes



1. Draw the major product of the reaction in the box above.

2. Circle the base below that can be used to force the equilibrium of the first step to the product side:

NaOH NH_3 (NaH)

3. Explain your answer in part b (why you chose one base **and** did not choose the others), using chemical structures as part of your answer.

Relevant pK_a values in H_2O:

NH_3 (NH_4^+): 9.2

NaOH (H_2O): 15.7

sp C–H: 24 (acetylene; estimated)

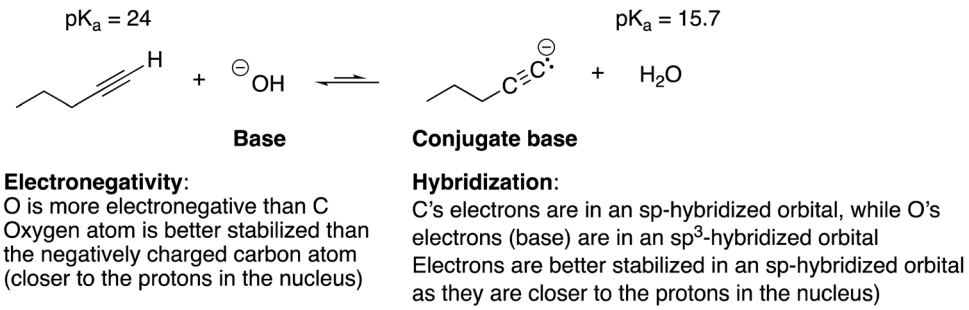
NaH (H_2): 35 (estimated, see *J. Chem. Soc., Chem. Comm.*, 1976, 648.)

Arguments (Table 6)

Expected (based on learning outcomes and expectations in that specific question): For one of the indicated bases to drive the acid–base equilibrium in the reaction illustrated above towards the products, it must be a less stable and stronger base than the carbanion. A base with a pK_a value of its conjugate acid greater than that of the alkyne (24) will be sufficiently strong. The pK_a values of the conjugate acids of NaH, NaOH, and NH_3 are 35, 15.7, and 9.2, respectively. Therefore, NaH is the only base capable of driving the equilibrium to products as it is the only base with a pK_a of its conjugate acid (35) greater than that of the alkyne (24). Using either NaOH or NH_3 would instead result in the equilibrium favouring reactants.

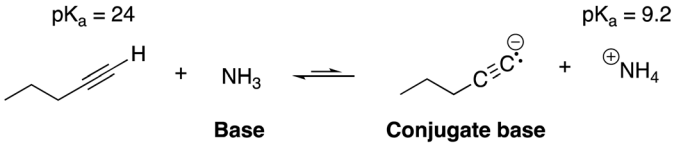
More granular causal argument (with chemical reasons as justification, molecular and atomic): Chemical properties and pK_a values together can be used to justify why NaOH and NH_3 will not be suitable bases to drive the acid–base equilibrium to products, and why H_2 will be suitable. These properties are used to explain the relative stability of the base (sodium hydroxide, ammonia, or sodium hydride) and conjugate base (acetylide).

The base and conjugate base involved in the equilibrium with sodium hydroxide have two competing chemical factors affecting their stability: the oxygen atom in the hydroxide (base) is more electronegative than the carbon atom in the conjugate base, which stabilizes the negative charge of the base more than the conjugate base. The electrons in the conjugate base are in an sp-hybridized orbital, which stabilizes the electrons more than the oxygen atom's electrons in the base that are in an sp^3-hybridized orbital (by virtue of being closer to the protons in the nucleus and therefore having a lesser effective negative charge). pK_a evidence indicates that the electronegativity factor dominates over the hybridization factor, with H_2O being a stronger acid (pK_a = 15.7) than the acetylene (pK_a = 24). The equilibrium favours the side with the weaker (more stable) species.



Ammonia will similarly not be a suitable base to drive the equilibrium to products. Hybridization factors stabilize the electrons on the C in the conjugate base more than on the N in the base (sp *versus* sp^3-hybrized, as above). However, both electronegativity and charge factors contribute to greater stability of the base than the conjugate base. The greater electronegativity of N (base) compared to C (conjugate base) stabilizes the base more than the conjugate base. Since charge decreases the stability of a species, the neutral nitrogen atom (base) is more stabilized than the negatively charged carbon atom (conjugate base). pK_a evidence indicates that the electronegativity and charge factors dominate over the hybridization factor, with ammonium being a stronger acid (pK_a = 9.2) than the acetylene (pK_a = 24). The equilibrium favours the side with the weaker (more stable) species.

 This journal is © The Royal Society of Chemistry 2021



Electronegativity:
N is more electronegative than C. Nitrogen atom is better stabilized than the negatively charged carbon atom (closer to the protons in the nucleus)

Charge:
Charge decreases the stability of a species, making the neutral N in the base more stable than the negatively charged C in the conjugate base.

Hybridization:
C's electrons are in an sp-hybridized orbital, while N's electr (base) are in an sp^3-hybridized orbital
Electrons are better stabilized in an sp-hybridized orbital as are closer to the protons in the nucleus)

pK$_a$ values (experimental evidence) indicate that electronegativity and charge dominate over hybridization in this case. The equilibrium favours the side with the weaker (more stable) species.

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Sodium hydride will be a suitable base to drive the equilibrium to products. Hybridization/Orbital factors stabilize the electrons on the H in the base more than on the C in the conjugate base (s *versus* sp-hybrized). However, both electronegativity and atom size contribute to greater stability of the conjugate base. The greater electronegativity of C (conjugate base) compared to H (base) stabilizes the conjugate base more than the base. Because carbon is a larger atom than hydrogen, the larger atom can better disperse the electron density (and negative charge), stabilizing the conjugate base more than the base. pK_a evidence indicates that the electronegativity and atom size factors dominate over the hybridization/orbital factor, with the acetylene being a stronger acid (pK_a = 25) than the acetylene (pK_a = 35). The equilibrium favours the side with the weaker (more stable) species, in this case, the product.

Appendix C: CHM 2120 2017 intended and enacted learning outcomes for Q1 and Q2

Acid–base learning outcomes

Required skills

• Draw the mechanism of an acid–base reaction, given the starting materials

• Identify the acid, base, conjugate acid, and conjugate base, given the starting materials

• Deprotonate a given molecule

• Protonate a given molecule

• Draw the conjugate acid and conjugate base, given a molecule

• Estimate the pK_a value of a given molecule

Key concepts

Apply the following ideas

• The stronger the acid, the weaker its conjugate base (and *vice versa*)

• An equilibrium favours the direction with the weaker (most stable) species (acid or base)

• The lower the pK_a value, the stronger the acid

• The following terms are synonymous: stronger = less stable = higher energy and weaker = more stable = lower energy

Applying required skills and key concepts to:

• Compare pK_a values of acids

• Compare relative stabilities of two species (*e.g.*, bases or acids), analyzing the effects of the following chemical principles: electronegativity, atom size, resonance, hybridization, inductive effects, charge, solvent

In the following contexts:

• Within a single molecule

• Between multiple molecules

• In an acid–base equilibrium

Integrate acid–base concepts in situations including:

• Identify the strongest acid and base that can exist in a given solvent

• Identify the predominant form of a molecule at a given pH (Henderson–Hasselbalch equation)

• Draw (or select from a list) a base/acid that could quantitatively deprotonate/protonate a given acid/base

Note: For associated explanations and activities, see http://www.flynnresearchgroup.com/acid-base and/or the acid–base module in https://orgchem101.com/ (Flynn).

Course learning outcomes relevant to comparing mechanisms:

• Identify the leaving group, α carbon, base, nucleophile, electrophile, and solvent in a given reaction

• Decide whether a reaction is likely to proceed *via* an E1/S$_N$1, E2, or S$_N$2 mechanism using structural and energetic features

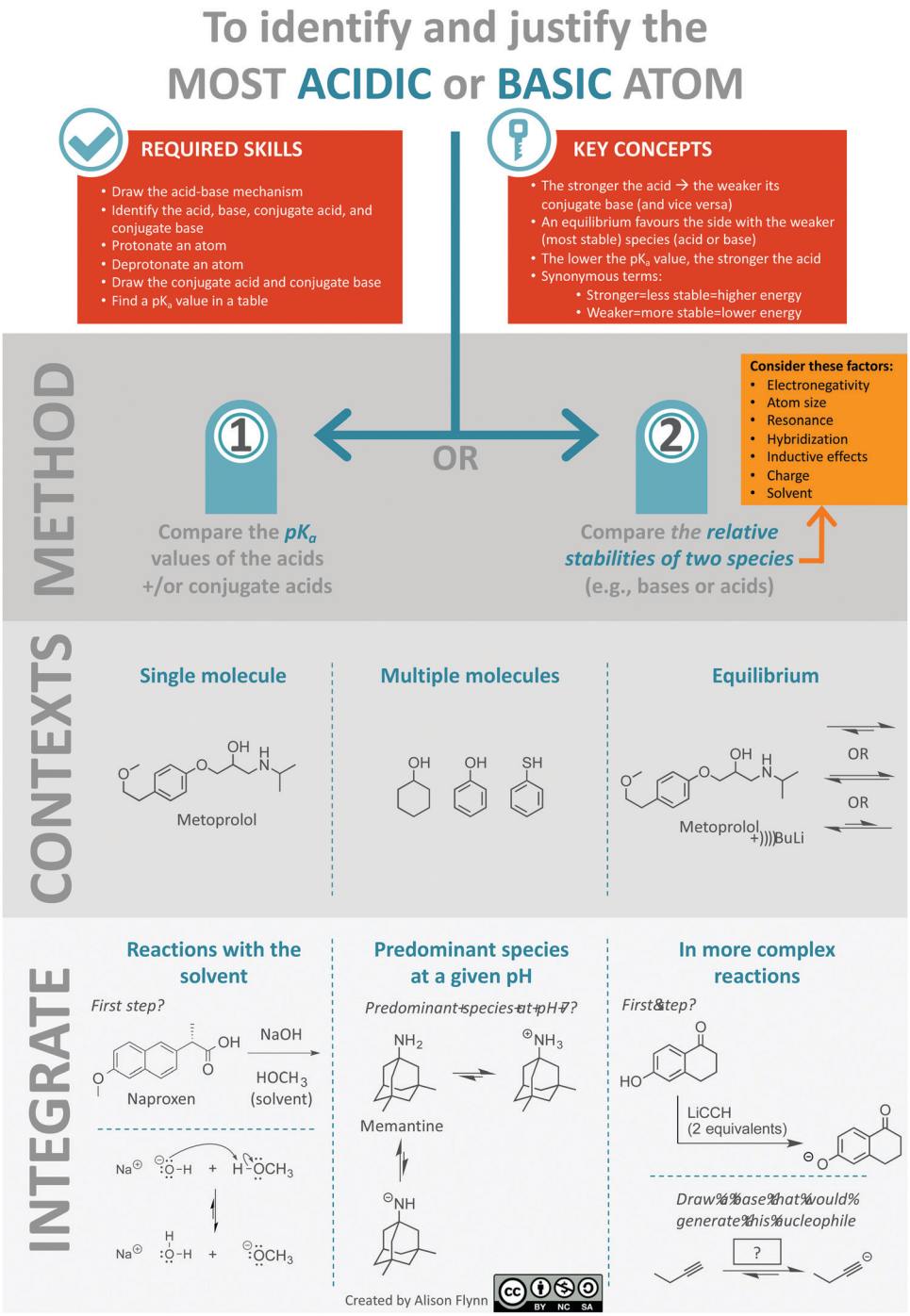
• Draw a mechanism for an E1, S$_N$1, E2, and S$_N$2 reaction, given various starting materials

• Identify appropriate solvents for each reaction

• Draw and label the reaction coordinate diagram for E1, S$_N$1, E2, and S$_N$2 reactions

• Predict the major product of the reaction, including stereochemical and regiochemical considerations

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.



Acknowledgements

We thank Myriam Carle for her assistance with the inter-rater reliability portion of this study. JD thanks the Natural Sciences and Engineering Research Council for funding in the form of a Canadian Graduate Scholarship (Master's).

Notes and references

Abrams E. and Southerland S., (2010), The how's and why's of biological change: How learners neglect physical mechanisms in their search for meaning, *Int. J. Sci. Educ.*, **23**(12), 1271–1281.

Banerjee A. C., (1991), Misconceptions of students and teachers in chemical equilibrium, *Int. J. Sci. Educ.*, **13**(4), 487–494.

Barwell R., (2018), Word problems as social texts, *Numer. Soc. Pract. Glob. Local Perspect.*, 101–120.

Berland L. K. and Reiser B. J., (2009), Making sense of argumentation and explanation, *Sci. Educ.*, **93**, 26–55.

Bernholt S. and Parchmann I., (2011), Assessing the complexity of students' knowledge in chemistry, *Chem. Educ. Res. Pract.*, **12**(2), 167–173.

 This journal is © The Royal Society of Chemistry 2021

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Bhattacharyya G., (2006), Practitioner development in organic chemistry: How graduate students conceptualize organic acids, *Chem. Educ. Res. Pract.*, **7**(4), 240–247.

Biggs J. and Tang C., (2011), Aligning assessment tasks with intended learning outcomes: principles, in *Teaching for Quality Learning at University*, pp. 191–223.

Bodé N. E., Deng J. M. and Flynn A. B., (2019), Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms, *J. Chem. Educ.*, **96**(6), 1068–1082.

Carle M. S. and Flynn A. B., (2020), Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chem. Educ. Res. Pract.*, **21**(2), 622–637.

Carle M. S., El Issa R., Pilote N. and Flynn A. B., (2020), Ten essential delocalization learning outcomes: How well are they achieved? *ChemRxiv*, 1–28.

Carmel J. H., Herrington D. G., Posey L. A., Ward J. S., Pollock A. M. and Cooper M. M., (2019), Helping Students to ''do Science'': Characterizing scientific practices in general chemistry laboratory curricula, *J. Chem. Educ.*, **96**(3), 423–434.

Cartrette D. P. and Mayo P. M., (2011), Students' understanding of acids/bases in organic chemistry contexts, *Chem. Educ. Res. Pract.*, **12**(1), 29–39.

Caspari I., Kranz D. and Graulich N., (2018a), Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework, *Chem. Educ. Res. Pract.*, **19**(4), 1117–1141.

Caspari I., Weinrich M. L., Sevian H. and Graulich N., (2018b), This mechanistic step is ''productive'': Organic chemistry students' backward-oriented reasoning, *Chem. Educ. Res. Pract.*, **19**(1), 42–59.

Cian H., (2020), The influence of context: Comparing high school students' socioscientific reasoning by socioscientific topic, *Int. J. Sci. Educ.*, **42**(9), 1–19.

Cooper M. and Klymkowsky M., (2013), Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform, *J. Chem. Educ.*, **90**(9), 1116–1122.

Cooper M. M., Kouyoumdjian H. and Underwood S. M., (2016), Investigating students' reasoning about acid–base reactions, *J. Chem. Educ.*, **93**(10), 1703–1712.

Crandell O. M., Kouyoumdjian H., Underwood S. M. and Cooper M. M., (2018), Reasoning about reactions in organic chemistry: Starting it in general chemistry.

Darden L., (2002), Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining, *Philos. Sci.*, **69**(S3), 354–365.

DeCocq V. and Bhattacharyya G., (2019), TMI (Too much information)! Effects of given information on organic chemistry students' approaches to solving mechanism tasks, *Chem. Educ. Res. Pract.*, **20**(1), 213–228.

Dixson D. D. and Worrell F. C., (2016), Formative and summative assessment in the classroom, *Theory Pract.*, **55**(2), 153–159.

Duis J. M., (2011), Organic chemistry educators' perspectives on fundamental concepts and misconceptions: An exploratory study, *J. Chem. Educ.*, **88**(3), 346–350.

Emig J., (1977), Writing as a mode of learning, *Coll. Compos. Commun.*, **28**(2), 122–128.

European Union, (2006), Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning, *Off. J. Eur. Union*, L394/19–L394/18.

Flynn A. B., (2017), Flipped chemistry courses: Structure, aligning learning outcomes, and evaluation, in *Online Approaches to Chemical Education*, American Chemical Society, pp. 151–164.

Flynn A. B., (2015), Structure and evaluation of flipped chemistry courses: Organic & spectroscopy, large and small, first to third year, English and French, *Chem. Educ. Res. Pract.*, **16**(2), 198–211.

Flynn A. B., *OrgChem101*.

Flynn A. B. and Amellal D. G., (2016), Chemical information literacy: pK_a values-where do students go wrong? *J. Chem. Educ.*, **93**(1), 39–45.

Flynn A. B. and Ogilvie W. W., (2015), Mechanisms before reactions: A mechanistic approach to the organic chemistry curriculum based on patterns of electron flow, *J. Chem. Educ.*, **92**(5), 803–810.

Grimberg B. I. and Hand B., (2009), Cognitive pathways: Analysis of students' written texts for science understanding, *Int. J. Sci. Educ.*, **31**(4), 503–521.

Hackling M. W. and Garnett P. J., (1985), Misconceptions of chemical equilibrium, *Eur. J. Sci. Educ.*, **7**(2), 205–214.

Hallgren K. A., (2012), Computing inter-rater reliability for observational data: An overview and tutorial, *Tutor. Quant. Methods. Psychol.*, **8**(1), 23–34.

Huddle P. A. and Pillay A. E., (1996), An in-depth study of misconceptions in stoichiometry and chemical equilibrium at a South African University, *J. Res. Sci. Teach.*, **33**(1), 65–77.

Jimenez-Aleixandre M. P. and Federico-Agraso M., (2009), Justification and persuasion about cloning: Arguments in Hwang's paper and journalistic reported versions, *Res. Sci. Educ.*, **39**(3), 331–347.

Jones M. D. and Crow D. A., (2017), How can we use the ''science of stories'' to produce persuasive scientific stories, *Palgrave Commun.*, **3**(1), 1–9.

Kelly G. J., Druker S. and Chen C., (1998), Students' reasoning about electricity: Combining performance assessments with argumentation analysis, *Int. J. Sci. Educ.*, **20**(7), 849–871.

Kraft A., Strickland A. M. and Bhattacharyya G., (2010), Reasonable reasoning: Multi-variate problem-solving in organic chemistry. *Chem. Educ. Res. Pract.*, **11**(4), 281–292.

Krajcik J. S. and Nakhleh M. B., (1994), Influence of levels of information as presented by different technologies on students' understanding of acid, base, and pH concepts, *J. Res. Sci. Teach.*, **31**(10), 1077–1096.

Krippendorff K., (1970), Estimating the reliability, systematic error and random error of interval data, *Educ. Psychol. Meas.*, **30**(1), 61–70.

Kuhn D., (2011), *The skills of argument*, Cambridge University Press.

Laverty J. T., Underwood S. M., Matz R. L., Posey L. A., Carmel J. H., Caballero M. D., et al., (2016), Characterizing college

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

science assessments: The three-dimensional learning assessment protocol, *PLoS One*, **11**(9), 1–21.

Luisi P. L., (2002), Emergence in chemistry: Chemistry as the embodiment of emergence, *Found. Chem.*, **4**(3), 183–200.

Machamer P., Darden L. and Craver C. F., (2000), Thinking about Mechanisms, *Philos. Sci.*, **67**(1), 1–25.

MacRie-Shuck M. and Talanquer V., (2020), Exploring Students' Explanations of Energy Transfer and Transformation, *J. Chem. Educ.*, **97**(12), 4225–4234.

Maeyer J. and Talanquer V., (2013), Making predictions about chemical reactivity: Assumptions and heuristics, *J. Res. Sci. Teach.*, **50**(6), 748–767.

McClary L. and Talanquer V., (2011), Heuristic reasoning in chemistry: Making decisions about acid strength, *Int. J. Sci. Educ.*, **33**(10), 1433–1454.

McNeill K. L., Lizotte D. J., Krajcik J. and Marx R. W., (2006), Supporting students' construction of scientific explanations by fading scaffolds in instructional materials, *J. Learn. Sci.*, **15**(2), 153–191.

Moon A., Moeller R., Gere A. R. and Shultz G. V., (2019), Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification, *Chem. Educ. Res. Pract.*, **20**(3), 484–494.

Moreira P., Marzabal A. and Talanquer V., (2019), Using a mechanistic framework to characterise chemistry students' reasoning in written explanations, *Chem. Educ. Res. Pract.*, **20**(1), 120–131.

National Research Council, (2012), *A Framework for K-12 Science Education*, National Academies Press.

Ogilvie W. W., Ackroyd N., Browning S., Deslongchamps G., Lee F. and Sauer E., (2017), *Organic Chemistry: Mechanistic Patterns*, 1st edn, Nelson Education Ltd.

Organisation for Economic Cooperation and Development, (2006), *Assessing scientific, reading and mathematical literacy: A framework for PISA*.

Orgill M. and Sutherland A., (2008), Undergraduate chemistry students' perceptions of and misconceptions about buffers and buffer problems, *Chem. Educ. Res. Pract.*, **9**(2), 131–143.

Osborne J. F. and Patterson A., (2011), Scientific argument and explanation: A necessary distinction? *Sci. Educ.*, **95**(4), 627–638.

Quilez-Pardo J. and Solaz-Portoles J. J., (1995), Students' and teachers' misapplication of Le Chatelier's Principle: Implications for the teaching of chemical equilibrium, *J. Res. Sci. Teach.*, **32**(9), 939–957.

Raycroft M. A. R. and Flynn A. B., (2020), What works? What's missing? An evaluation model for science curricula that analyses learning outcomes through five lenses, *Chem. Educ. Res. Pract.*, **21**(4), 1110–1131.

Reed J. J., Brandriet A. R. and Holme T. A., (2017), Analyzing the role of science practices in ACS exam items, *J. Chem. Educ.*, **94**(1), 3–10.

Sadler T. D., (2004), Informal reasoning regarding socioscientific issues: A critical review of research, *J. Res. Sci. Teach.*, **41**(5), 513–536.

Sadler T. D. and Zeidler D. L., (2005), The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues, *Sci. Educ.*, **89**(1), 71–93.

Sevian H. and Talanquer V., (2014), Rethinking chemistry: a learning progression on chemical thinking, *Chem. Educ. Res. Pr.*, **15**(1), 10–23.

Sevian H., Bernholt S., Szteinberg G. A. and Auguste S., (2015), Use of representation mapping to capture abstraction in problem solving in different courses in chemistry, *Chem. Educ. Res. Pract.*, **16**(3), 429–446.

Social Sciences and Humanities Research Council, (2018), *Truth Under Fire in a Post-Fact World*.

Southard K. M., Espindola M. R., Zaepfel S. D. and Molly S., (2017), Generative mechanistic explanation building in undergraduate molecular and cellular biology, *Int. J. Sci. Educ.*, **39**(13), 1795–1829.

Stowe R. L. and Cooper M. M., (2017), Practicing what we preach: Assessing ''critical thinking'' in organic chemistry, *J. Chem. Educ.*, **94**(12), 1852–1859.

Stoyanovich C., Gandhi A. and Flynn A. B., (2015), Acid–base learning outcomes for students in an introductory organic chemistry course, *J. Chem. Educ.*, **92**(2), 220–229.

Talanquer V., (2007), Explanations and teleology in chemistry education, *Int. J. Sci. Educ.*, **29**(7), 853–870.

Talanquer V., (2017), Concept inventories: Predicting the wrong answer may boost performance, *J. Chem. Educ.*, **94**(12), 1805–1810.

Talanquer V., (2018a), Assessing for chemical thinking, in *Research and Practice in Chemistry Education*, Springer Nature Singapore Pte Ltd, pp. 123–133.

Talanquer V., (2018b), Progressions in reasoning about structure – property relationships, *Chem. Educ. Res. Pract.*, **19**(4), 998–1009.

Talanquer V. and Pollard J., (2010), Let's teach how we think instead of what we know, *Chem. Educ. Res. Pract.*, **11**(2), 74–83.

Toulmin S., (1958), *The Uses of Argument*, Cambridge University Press.

Trommler F., Gresch H., Hammann M., Trommler F., Gresch H. and Hammann M., (2018), Students' reasons for preferring teleological explanations, *Int. J. Sci. Educ.*, **40**(2), 159–187.

United Nations, (2015), *Transforming our World: the 2030 Agenda for Sustainable Development*.

Verheij B., (2003), Dialectical argumentation with argumentation schemes: An approach to legal logic, *Artif. Intell. Law*, **11**(2–3), 167–195.

van Mil M. H. W., Jan D., Arend B. and Waarlo J., (2013), Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour, *Sci. Educ.*, **22**(1), 93–118.

von Aufschnaiter C., Erduran S., Osborne J., Simon S., Education P. and Giessen J., (2008), Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge, *J. Res. Sci. Teach.*, **45**(1), 101–131.

Voska K. W. and Heikkinen H. W., (2000), Identification and analysis of student conceptions used to solve chemical equilibrium problems, *J. Res. Sci. Teach.*, **37**(2), 160–176.

 This journal is © The Royal Society of Chemistry 2021

Weinrich M. L. and Sevian H., (2017), Capturing students' abstraction while solving organic reaction mechanism problems across a semester, *Chem. Educ. Res. Pract.*, **18**(1), 169–190.

Weinrich M. L. and Talanquer V., (2016), Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product, *Chem. Educ. Res. Pract.*, **17**(2), 394–406.

Wheeler A. E. and Kass H., (1978), Student misconceptions in chemical equilibrium, *Sci. Educ.*, **62**(2), 223–232.

Windschitl M., Thompson J. and Braaten M., (2008), Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations, *Sci. Educ.*, **92**(5), 941–967.

Published on 16 April 2021. Downloaded on 4/18/2024 10:31:04 PM.

Chapter 4: Students' Reasoning in Chemistry Arguments and Designing Resources Using Constructive Alignment

4.1 Introduction to the research article presented in this chapter

Chapter 4 is a book chapter I co-authored with Dr. Myriam McKenna, a former Flynn group PhD student, and Dr. Flynn. The chapter draws on my research described in Chapter 2 and 3, combining our findings with Dr. McKenna's expertise on constructive alignment, learning outcomes, and course design. Argumentation and reasoning are essential skills for citizens in a rapidly changing world, and educators need to be equipped with strategies and approaches to support students' development of these skills. Guided by this need, our book chapter provides suggestions on how to support students' reasoning and argumentation in chemistry through strategic course design (*e.g.*, backwards design, constructive alignment). Dr. McKenna (formerly Carle) and I are co-first authors on this chapter.

4.2 Copyright

Reproduced from "Deng, J. M.; Carle, M. S.; Flynn, A. B. Students' Reasoning in Chemistry Arguments and Designing Resources Using Constructive Alignment. In *Student Reasoning in Organic Chemistry*; Shultz, G. V., Graulich, N., Eds.; Royal Society of Chemistry, 2022. doi.org/10.1039/9781839167782-00074" with permission from the Royal Society of Chemistry. Deng and Carle are co-first authors.

CHAPTER 5

Students' Reasoning in Chemistry Arguments and Designing Resources Using Constructive Alignment

JACKY M. DENG[†a], MYRIAM S. CARLE[†a] AND ALISON B. FLYNN*[a]

[a]University of Ottawa, 10 Marie Curie Pvt, Ottawa, K1N 9A4, Canada
*E-mail: alison.flynn@uottawa.ca

5.1 Introduction

5.1.1 Citizens Need to be Able to Reason with Scientific Evidence

In a world facing complex global issues, citizens need to be able to make and justify claims informed by scientific evidence and reasoning.[1,2] National frameworks for science education around the world have also identified that constructing evidence-based arguments is a key scientific practice.[2–5]

Arguments justify a fact or phenomenon that is not agreed-upon;[6,7] a claim is in doubt and must be advanced through reasoning by constructing an argument about the fit between the evidence and claim.[8–10] For example,

[†]Co-lead authors.

Advances in Chemistry Education Series No. 10
Student Reasoning in Organic Chemistry: Research Advances and Evidence-based Instructional Practices
Edited by Nicole Graulich and Ginger Shultz
© The Royal Society of Chemistry 2023
Published by the Royal Society of Chemistry, www.rsc.org

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

making a decision (claim) of whether to vaccinate requires choosing sources of evidence, interpreting the quality of the available evidence, and using this evidence to reason for or against a particular decision.[11] In science and chemistry education, students' written and oral arguments about phenomena have been used to gain insight into students' reasoning.[8,10,12–22] Reasoning has been the focus of many chemistry education research investigations analyzing students' arguments and other contexts. There are several reasoning frameworks,[23–38] and each has been used to study different ways of thinking about students' approaches to problems and phenomena in chemistry and other fields.

Our research in student reasoning arose from anecdotal observations in exam responses, in which some students often stated evidence alone as justification for a given phenomenon, without going further to explain the relevance or impacts of that evidence (the "who cares"). While justifying phenomena (scientific argumentation) had been a long-time expectation in our department, a new curriculum that focused on patterns and principles of reactivity made scaffolding students' scientific argumentation skill development even more of a priority.[39,40]

5.2 Framework—Reasoning, Granularity, and Comparisons

We developed an analytical framework that characterizes arguments in terms of three dimensions: reasoning, granularity, and comparisons (Figure 5.1), which we describe below. In the chemistry education literature, students' arguments have primarily been analyzed with a focus on students' reasoning. However, characterizing students' reasoning does not necessarily capture other dimensions of an argument that may be of interest to chemistry

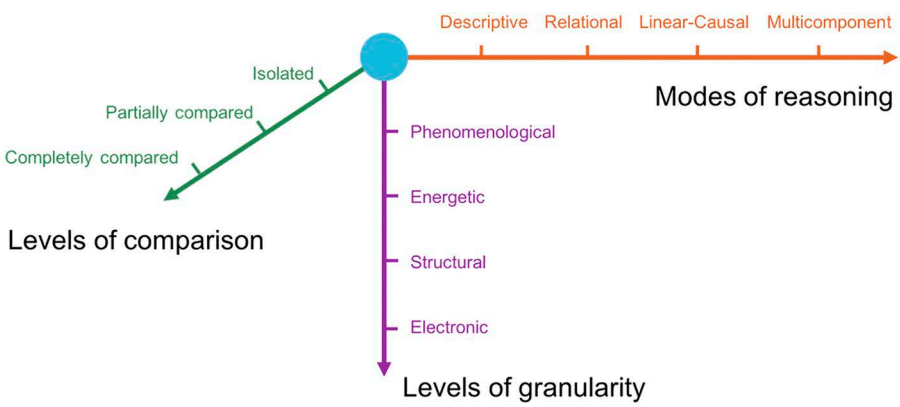


Figure 5.1 Overview of the framework: levels of comparison, granularity, and mode of reasoning. Reproduced from ref. 21 with permission from the Royal Society of Chemistry.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

educators and researchers, such as discussions at various scalar levels and/or comparisons against alternative claims.

5.2.1 Modes of Reasoning

We analyzed students' arguments using four modes of reasoning: descriptive, relational, linear causal, and multi-component causal (Table 5.1).[30,41] *Descriptive* arguments list or give features and/or the properties of entities (*e.g.*, the reactants, products) without establishing connections. *Relational* arguments include connections between properties of the entities and their activities, but these relationships are discussed in a correlative fashion (*i.e.*, lack causality). *Causal* arguments—linear and multi-component—include all features of a relational argument and additionally contain cause-and-effect relationships between the relevant properties of the entities and their activities. The modes have been used across various chemistry contexts to characterize students' reasoning in written and verbal argumentation.[12,20,21,35,42,43]

The type(s) of reasoning needed can vary depending on the task/context. In some contexts, causal reasoning may be helpful, while in others, descriptive or relational reasoning may be more practical and appropriate for accomplishing a task in educational and scientific settings. For example, how students choose to determine the direction of an acid–base equilibrium may differ on a final exam compared to a lab context. On a final exam, students may choose to construct an argument with causal reasoning to demonstrate their understanding (comparing pK_a values and discussing why the pK_a values are different based on chemical factors). In contrast, students may instead demonstrate only descriptive or relational reasoning (comparing pK_a values) to accomplish a similar task in the lab.

5.2.2 Levels of Granularity—Moving Between Grain Sizes

In addition to different modes of reasoning, arguments can also be constructed at different levels of granularity. For instance, justifications for why acetylsalicylic acid acts as an acid in water may focus on pH and pK_a data (*i.e.*, a phenomenological level) or how the stability of the deprotonated form of acetylsalicylic acid is stabilized by delocalization (*i.e.*, structural, electronic, and energetic factors).[44] Both justifications discuss a cause-and-effect but are constructed at different "scales" or levels of granularity.[45] The idea of granularity has been described in other work related to scientific reasoning, including scales,[46] levels,[47] nested hierarchies,[48] emergence (with ideas of downward and upward causality),[49] and bottom-out reasoning.[50]

Chemists need to be able to move between different levels of granularity.[49,51,52] Scientists may discuss how interactions between components at a submicroscopic level manifest into macroscopic properties.[49,51,52] Different contexts and tasks may require different levels of granularity, as phenomena can be explained from larger perspectives (*e.g.*, global levels and beyond) or smaller perspectives (*e.g.*, atomic levels and beyond).[50]

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

Table 5.1 Modes of reasoning. Reproduced from ref. 21 with permission from the Royal Society of Chemistry.

Mode of reasoning	Description	Example
Descriptive	Claims are supported solely by describing features and/or the properties of entities (*e.g.*, the reactants, products, *etc.*). No connections are established.	NaH is a strong base so it will drive the equilibrium to products.
Relational	Relationships between the properties of the reaction materials and their activities are discussed in a correlative fashion (*i.e.*, a cause–effect relationship is not established).	NaH will drive the equilibrium towards products because its conjugate acid has a higher pK_a value (35) than the alkyne (25).
Linear Causal	All the features of a relational argument are present and are accompanied by cause-and-effect relationships between the relevant properties of the entities and their activities.	NaH will drive the equilibrium towards products because its conjugate acid (pK_a value 35) is weaker/more stable than the starting alkyne (pK_a value 25), according to the relevant pK_a values. NaOH and NH_3 are not strong enough bases to drive the equilibrium to the product side, as both of their conjugate acids are stronger than the starting alkyne (pK_a values of 15.7 and 9, respectively).
Multi-Component Causal	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct interactions of several components.	The oxygen atom in NaOH is more electronegative than the carbon atom in the conjugate base, making it better able to stabilize the negative charge. Although atomic size (larger carbon atom) and hybridization suggest that the carbon atom in the conjugate base could be more stable, the pK_a values of the acid and conjugate acids support the conclusion that NaOH is the weaker base, since the stronger the acid, the weaker its conjugate base (pK_a value of H_2O = 15.7; pK_a of alkyne = 25). NH_3 is similarly too weak to drive the equilibrium to the product side, with the nitrogen atom being neutral in addition to being more electronegative than the carbon anion.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

In our research, we characterized the granularity exhibited in students' organic chemistry arguments in terms of four levels of granularity developed from literature: phenomenological, structural, electronic, and energetic.[21,44,47–49,53]

5.2.3 Comparison—Considering Alternatives

A comparison is needed when an argument involves two or more possible claims since reasoning and granularity do not capture how one considers alternatives within an argument.[9] Comparing alternatives is key to argumentation—when deciding on a claim, one must consider multiple possibilities and justify their claim by weighing various factors, evidence, *etc.* For example, justifying why phenol is a stronger acid than ethanol may include a description of why phenolate *can* delocalize and why ethoxide *cannot*. Arguments may include full comparisons (comparing choices by considering all factors), partial comparisons (comparing choices by considering some, but not all factors), or no comparisons (not comparing at all).

5.3 Students' Arguments Can Vary Between Tasks

We characterized the reasoning, granularity, and comparisons demonstrated in students' written arguments on three organic chemistry final exam items: (1) comparing two reaction mechanisms and justifying which was more plausible ($N = 159$);[20] (2) comparing three bases and justifying which will drive an equilibrium to products ($N = 170$);[21] and (3) determining resonance structures' relative contributions to the respective resonance hybrid ($N = 284$)[43] (Figure 5.2). Across the three questions, we found that students' responses varied in terms of their reasoning, granularity, and comparisons (Figure 5.3).

Educators' expectations about students' arguments (based on intended and enacted course learning outcomes (LOs)) were different *between* each question we analyzed. However, *within* the subset of responses for each question, students' arguments aligned with the expected reasoning level.[20,21,43] In the question comparing mechanisms (Q1), students had been taught to compare the structure of the starting material and explain how and why the structural features informed reaction plausibility, but the answer could be obtained using a simple rule-of-thumb (*e.g.*, relative carbocation stability guides reactivity, with methyl $\ll$ primary $<$ secondary $<$ tertiary). The educator explicitly explained and modeled that the rule was not enough to answer the question and that students had to explain why the rule was relevant. In other words, students' reasoning was aligned with how they were taught and with the course expectations. In the acid–base question (Q2), the question could be answered with two different levels of granularity. Either was expected and accepted: energetic and electronic factors (*e.g.*, formal charge, electronegativity) or phenomenological factors (*e.g.*, pK_a values); answers were distributed as such for this question. Similarly,

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

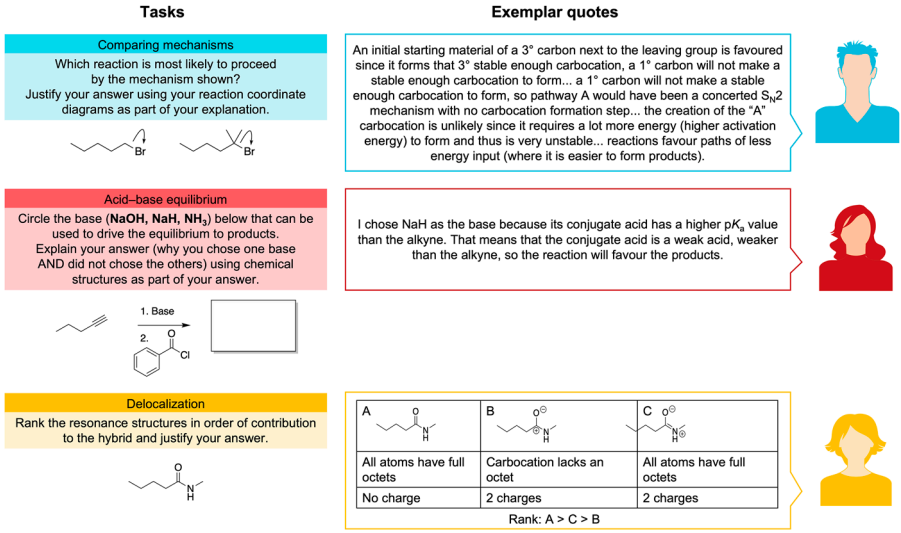


Figure 5.2 The three questions analyzed and exemplar quotes from students that reflect the expected arguments for each task.

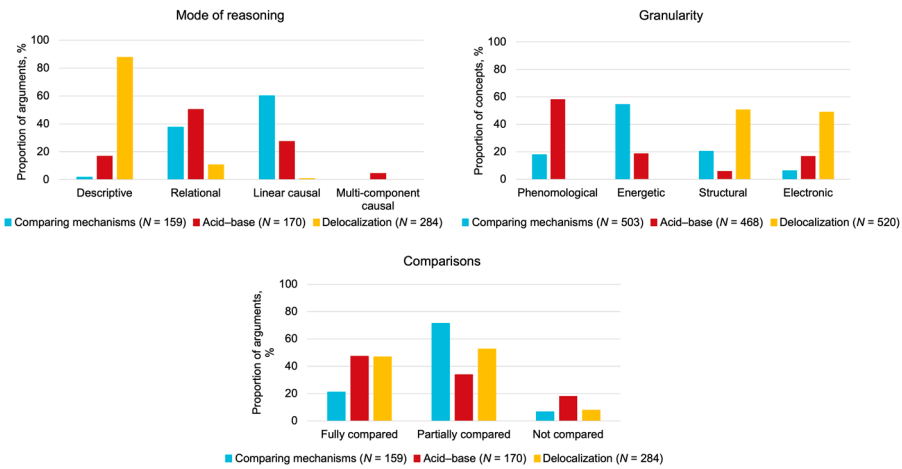


Figure 5.3 Summary of the reasoning, granularity, and comparisons demonstrated in students' arguments for the three questions.

for the delocalization question (Q3), the students used the same reasoning they were taught, aligned with expectations.

The arguments we analyzed demonstrated that students *could* produce sophisticated justifications of chemical phenomena in some tasks while producing predominantly descriptive arguments in other tasks. The structure of students' arguments may therefore be associated with various factors, such as the task's expectations, the task's structure, LOs, student knowledge, and

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

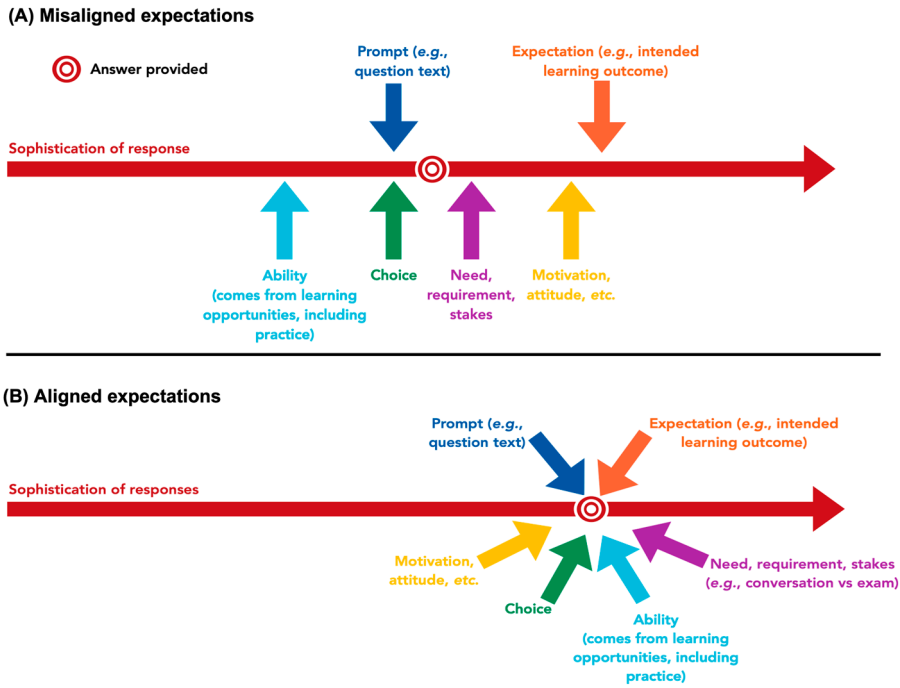


Figure 5.4 Students' responses along with (A) misaligned factors, and (B) aligned factors.

student motivation. Each of these factors should be carefully communicated throughout a course, to make expectations clear and best support students' learning. Misalignment of these factors with an expected answer (Figure 5.4, top) could lead to student confusion about expectations and result in *answers* that do not meet the expectations, despite students' *abilities* to meet them. With effective alignment of the factors, students have a higher chance of knowing the expectations and being able to provide the expected answer (Figure 5.4, bottom). In the following sections, we will describe how readers can use constructive alignment to help students develop and demonstrate their reasoning, granularity, and comparisons within arguments.

5.4 Supporting Student Learning Through Constructive Alignment

5.4.1 Instructional Design

Ideally, the learning environment is inclusively designed in a way that aligns the intended LOs with instruction and assessment—this is known as constructive alignment[54,55] and represents a shift in focus from what

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

the educator *does* to supporting what learners *need*. Viewed through the lens of constructive alignment, supporting students' explanations of chemical phenomena requires not only changing instructional practices, but also aligning the LOs, assessment, pedagogy, and other instructional decisions.[56,57]

One way to facilitate constructive alignment is through backwards design[58] in which educators design activities, assessments, and a learning environment by: (1) identifying the intended LOs, (2) developing practice opportunities and identifying acceptable evidence of achievement (*i.e.*, design of formative and summative assessments, including scaffolds), and (3) planning the learning experience (*i.e.*, pedagogy and other instructional decisions).

A general LO could take the form: justify the reactivity for [reactivity learned in Module x] by comparing the relevant entities and activities at [level of granularity] using [descriptive, relational, linear causal, or multi-component causal] reasoning (Table 5.2). An LO like this can be used to clearly communicate expectations and provided as metacognitive tool for students.[59] Such LOs can also be communicated in multiple ways: in a course syllabus, at the beginning/end of each lecture, class, *etc.*

Once the LOs are established, *we can analyze how they are enacted* by looking at how they are taught, practiced and assessed.[40,56,60] Assessment is how we can determine whether students have demonstrated acceptable achievement of LOs—we ask: *what evidence would demonstrate that a student has achieved the intended LOs?* From this, educators can begin to make decisions about the type(s) of assessment needed to adequately assess students' achievement of the LOs. Similarly, students can use the LOs as a tool to create their own mock practice questions.

Table 5.2 The prompts use the LOs and explicitly demonstrate the expectation from each component of the reasoning framework.

Question	Revised prompt
Q1: Comparing mechanisms	Which reaction is most likely to proceed by the mechanism shown? *Justify your answer with multi-component causal reasoning* by ***comparing*** the two reaction coordinate diagrams using an explanation at the **molecular level.**
Q2: Acid–base	Circle the base below that can be used to drive the equilibrium to products. *Justify your answer using at least linear causal reasoning* at the **reaction or molecular level of granularity** (***why you chose one base and did not chose the others***) using chemical structures as part of your answer.
Q3: Delocalization	Rank the resonance structures in order of contribution to the hybrid. *Justify your answer at the descriptive level* by ***comparing*** the resonance ***structures***.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

Lastly, with the LOs, assessments, and expectations in hand, instruction can be designed to explicitly focus on the desired outcome. If educators expect learners to achieve a specific LO, explicitly teaching it and assessing it throughout a course sequence provides learners with the tools, opportunities, and knowledge needed to succeed.[56] Followed in this manner, backwards design that starts with LOs can ensure that students have the necessary learning opportunities throughout a course to successfully develop and demonstrate each LO.

5.4.2 Scaffolding Skill Development

Students' reasoning skill development can also be supported through scaffolds, supports, and intermediate steps that guide students toward the intended LO.[61] In the context of argumentation, scaffolds can provide the structure for students to practice constructing arguments with specific components in mind (*e.g.*, scaffolds that explicitly ask for claim, evidence, and reasoning). In accordance with constructive alignment, LOs and expectations need to guide the design of the scaffolds to build students' skills. For example, researchers have proposed ways to design scaffolds for organic chemistry that focus on outcomes related to mechanistic reasoning.[29,62] Another example could be using the fine-grained LOs seen in Table 5.2 as scaffolded steps. For the acid–base question analyzed in our research, a scaffold could first ask the students to provide the pK_a values of the protonated species (granularity), compare the three pieces of evidence (comparison), explain why pK_a is used in this case (causal reasoning) and finally state a claim. This type of scaffolding lays out the expectations for the learner, ensuring that they are aware of what they must demonstrate within their argument. Additionally, prompting students to think about what they have learned from scaffolds and to apply their knowledge to new tasks could help them develop metacognitive and transfer skills.[62,63] As time goes by, and students' skills and learning expectations increase, scaffolds can also be removed or faded, though research is needed on the timing and impacts of fading scaffolds on students' arguments in chemistry.[7]

The following example of scaffolding comes from an online organic chemistry course with 300 students on Zoom. The LO for the class was *Identify the claim, evidence, and reasoning* within a pre-written chemistry argument. Before class, students watched videos on scientific argumentation and reasoning, including modes of reasoning, comparisons, and granularity.[64,65] They also completed a pre-class quiz to assess their basic comprehension of the videos. In class, students first answered a multiple-choice question using an online response system: "Which substrate is most likely to react by the first step indicated, **A** or **B**?" (Figure 5.5). Next, they used the response system to give their reasoning, being prompted to provide a linear causal response (examples in Figure 5.5b). They were then asked to join a breakout room and identify the claim, evidence, and reasoning in one

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

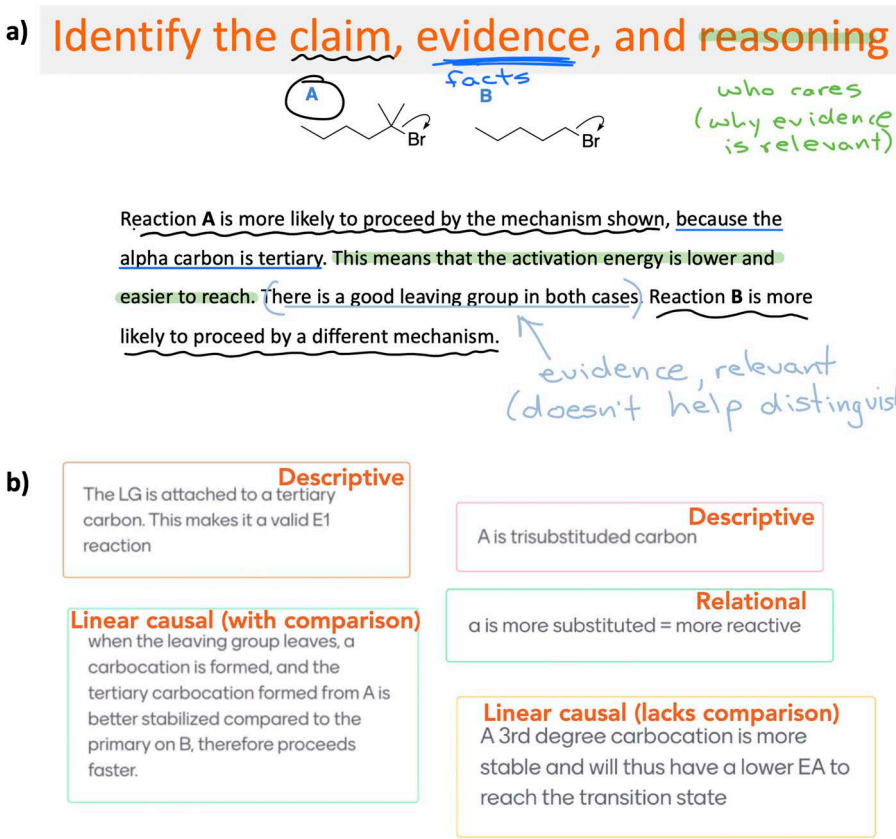


Figure 5.5 Students identified which substrate (A or B) was most likely to react by the first step shown. (a) Students worked together in breakout rooms to identify the claim, evidence, and reasoning in a pre-constructed argument. (b) Students identified the mode of reasoning in class examples, which had been written during the class.

answer (prepared in advance); students shared their answers back with the full class with an accompanying discussion (Figure 5.5a). Students were then shown a selection of their classmates' responses and asked to classify each one as descriptive, relational, linear causal, or multi-component causal (Figure 5.5b). They were prompted to revise and resubmit their own answers based on the discussion. Finally, the class moved on to a new example to further practice these skills.

Rubrics are another way to communicate expectations. Rubrics can be used to make explicit the expectations for a particular learning outcome—the rubric shown in Figure 5.6 could be used to guide how an instructor builds assessments and/or the teaching experience. The rubric could also be provided directly to students to show what is expected and how they will be assessed.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

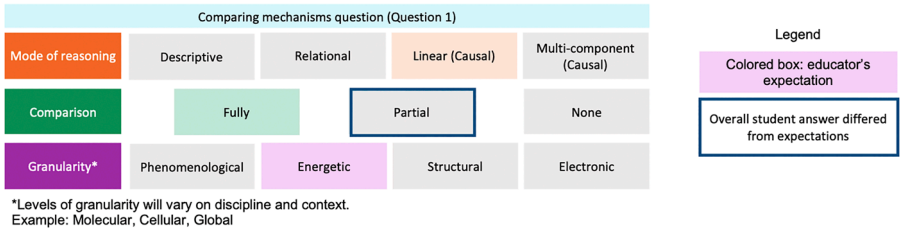


Figure 5.6 A rubric for the reasoning framework with the expectation of each question and alignment of answers with expectations.

5.4.3 Resources for Constructively Aligning Reasoning into a Course

The Flynn Research Group website[66] contains resources on how to use the reasoning, granularity, and comparisons framework and align it within a course.

5.5 Conclusions

In a world facing complex global issues, citizens need to be able to make and justify decisions, an important aspect of scientific argumentation skills. As chemistry educators, we can embed scientific argumentation into our teaching and assessment to support students' argumentation skills. Herein, we share a framework that incorporates three essential aspects to effective argumentation: reasoning (*i.e.*, descriptive, relational, linear causal, and multicomponent causal), granularity (the scalar levels discussed in arguments), and comparisons (*e.g.*, comparisons between possibilities, counterclaims).

From our research using the framework, we found that the structure of students' arguments varied between questions. For each question, students reasoning, granularity, and comparisons varied *between* tasks, but was consistent with the expectations associated *with* each task.

Therefore, we share how constructive alignment might be used to maximize students' opportunities to develop and demonstrate argumentation-related learning outcomes. With this approach, intended learning outcomes are used to guide instructional decisions, including pedagogy and assessment.

Future work could investigate how students perceive course expectations in more and less aligned curricula, what specific features of a course students rely on when making trying to make sense of expectations (in both aligned and misaligned curricula), how students' arguments develop over time in different curricula, the timing and effectiveness of fading scaffolds on student argumentation in chemistry, the effects of context on students' arguments,[51,67,68] and the effects of grading practices on students' skill development.[69] Argumentation and reasoning are essential skills for both scientists and citizens in a world facing complex challenges; researching

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

evidence-based educational practices, including constructive alignment, may be effective supports for students in developing and demonstrating these skills.

References

1. United Nations, Transforming our World: the 2030 Agenda for Sustainable Development [Internet], 2015. Available from: https://www.un.org/sustainabledevelopment/sustainable-development-goals/.
2. Social Sciences and Humanities Research Council, Truth Under Fire in a Post-Fact World [Internet], 2018 [cited 1 September 2019]. Available from: https://horizons.gc.ca/en/2018/10/19/the-next-generation-of-emerging-global-challenges/#truth-under-fire.
3. National Research Council, A Framework for K-12 Science Education [Internet], National Academies Press, Washington, D.C., 2012, Available from: https://www.nap.edu/read/13165/chapter/1.
4. European Union, Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning, *Off. J. Eur. Union*, 2006, L 394/19-L 394/18.
5. Organisation for Economic Cooperation and Development, *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*, 2006.
6. D. Kuhn, *The Skills of Argument*, New York, New York, USA, Cambridge University Press, 2011.
7. K. L. McNeill, D. J. Lizotte, J. Krajcik and R. W. Marx, Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials, *J. Learn. Sci.*, 2006 Apr, **15**(2), 153–191.
8. J. F. Osborne and A. Patterson, Scientific Argument and Explanation: A Necessary Distinction? *Sci. Educ.*, 2011, **95**(4), 627–638.
9. S. Toulmin, *The Uses of Argument*, Cambridge University Press, Cambridge, 1958.
10. K. L. McNeill and J. Krajcik, Inquiry and Scientific Explanations: Helping Students Use Evidence and Reasoning, in *Science as Inquiry in the Secondary Setting*, 2008, pp. 121–34.
11. M. D. Jones and D. A. Crow, How can we use the "science of stories" to produce persuasive scientific stories, *Palgrave Commun.*, 2017, **3**(1), 1–9.
12. A. Moon, C. Stanford, R. Cole and M. H. Towns, The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry, *Chem. Educ. Res. Pract.*, 2016, **17**(2), 353–364.
13. K. L. McNeill and L. K. Berland, A Learning Progression for Scientific Argumentation: Understanding Student Work and Designing Supportive, *Sci. Educ.*, 2010, 765–793.
14. V. Dawson and K. Carson, Using climate change scenarios to assess high school students' argumentation skills, *Res. Sci. Technol. Educ.*, 2017, **35**(1), 1–16.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

15. L. K. Berland and B. J. Reiser, Making Sense of Argumentation and Explanation, *Sci. Educ.*, 2009, **93**, 26–55.
16. J. Emig, Writing as a Mode of Learning, *Coll. Compos. Commun.*, 1977, **28**(2), 122–128.
17. B. I. Grimberg and B. Hand, Cognitive pathways: Analysis of students' written texts for science understanding, *Int J Sci Educ*, 2009, **31**(4), 503–521.
18. G. J. Kelly, J. Regev and W. Prothero, Analysis of Lines of Reasoning in Written Argumentation, in *Argumentation in Science Education*, ed. S. Erduran and M. P. Jimenez-Aleixandre, Springer Science, 2007, pp. 137–58.
19. N. Y. P. Diana, Assessment of Argumentation in Chemistry: A Model for Designing Items, in *Argumentation in Chemistry Education: Research, Policy and Practice*, 2019, pp. 106–41.
20. N. E. Bodé, J. M. Deng and A. B. Flynn, Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms, *J. Chem. Educ.*, 2019, **96**(6), 1068–1082.
21. J. M. Deng and A. B. Flynn, Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items, *Chem. Educ. Res. Pract.*, 2021, **22**, 749–771.
22. A. Moon, R. Moeller, A. R. Gere and G. V. Shultz, Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification, *Chem. Educ. Res. Pract.*, 2019, **20**(3), 484–494.
23. L. McClary and V. Talanquer, Heuristic reasoning in chemistry: making decisions about acid strength, *Int. J. Sci. Educ.*, 2011 Jul, **33**(10), 1433–1454.
24. V. Talanquer, Concept Inventories: Predicting the Wrong Answer May Boost Performance, *J. Chem. Educ.*, 2017, **94**(12), 1805–1810.
25. A. Kraft, A. M. Strickland and G. Bhattacharyya, Reasonable reasoning: multi-variate problem-solving in organic chemistry, *Chem. Educ. Res. Pract.*, 2010, **11**(4), 281–292.
26. G. Bhattacharyya and G. M. Bodner, "It gets me to the product": How students propose organic mechanisms, *J. Chem. Educ.*, 2005, **82**(9), 1402–1407.
27. J. L. Kolodner, An Introduction to Case-Based Reasoning, *Artif. Intell. Rev.*, 1992, **6**, 3–34.
28. N. J. Nersessian, Model-Based Reasoning in Conceptual Change, in *Model-Based Reasoning in Scientific Discovery*, 1999, pp. 5–22.
29. M. M. Cooper, H. Kouyoumdjian and S. M. Underwood, Investigating Students' Reasoning about Acid-Base Reactions, *J. Chem. Educ.*, 2016, **93**(10), 1703–1712.
30. R. S. Russ, R. E. Scherr, D. Hammer and J. Mikeska, Recognizing Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed From Philosophy of Science, *Sci. Educ.*, 2008, **92**(3), 499–525.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

31. J. Maeyer and V. Talanquer, Making Predictions About Chemical Reactivity: Assumptions and Heuristics, *J. Res. Sci. Teach.*, 2013, **50**(6), 748–767.
32. K. E. Stanovich and R. F. West, Individual differences in reasoning: Implications for the rationality debate? *Behav. Brain Sci.*, 2003, **26**(4), 527.
33. V. Talanquer, Explanations and Teleology in Chemistry Education, *Int. J. Sci. Educ.*, 2007, **29**(7), 853–870.
34. V. DeCocq and G. Bhattacharyya, TMI (Too much information)! Effects of given information on organic chemistry students' approaches to solving mechanism tasks, *Chem. Educ. Res. Pract.*, 2019, **20**(1), 213–228.
35. I. Caspari, D. Kranz and N. Graulich, Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework, *Chem. Educ. Res. Pr.*, 2018, **19**, 1117–1141.
36. L. Wright, Explanation and teleology, *Philos. Sci.*, 1972, **39**, 204–218.
37. M. L. Weinrich and H. Sevian, Capturing students' abstraction while solving organic reaction mechanism problems across a semester, *Chem. Educ. Res. Pract.*, 2017, **18**(1), 169–190.
38. D. L. Schwartz, The Emergence of Abstract Representations in Dyad Problem Solving, *J. Learn. Sci.*, 1995, **4**(3), 321–354.
39. A. B. Flynn and W. W. Ogilvie, Mechanisms before Reactions: A Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow, *J. Chem. Educ.*, 2015, **92**(5), 803–810.
40. M. A. R. Raycroft and A. B. Flynn, What works? What's missing? An evaluation model for science curricula that analyses learning outcomes through five lenses, *Chem. Educ. Res. Pract.*, 2020, **21**(4), 1110–1131.
41. H. Sevian and V. Talanquer, Rethinking chemistry: a learning progression on chemical thinking, *Chem. Educ. Res. Pract.*, 2014, **15**(1), 10–23.
42. P. Moreira, A. Marzabal and V. Talanquer, Using a mechanistic framework to characterise chemistry students' reasoning in written explanations, *Chem. Educ. Res. Pract.*, 2019, **20**(1), 120–131.
43. M. S. Carle, R. El Issa, N. Pilote and A. B. Flynn, Ten essential delocalization learning outcomes: How well are they achieved? ChemRxiv, 2021, 1–28.
44. V. Talanquer, Assessing for Chemical Thinking, in *Research and Practice in Chemistry Education*, Springer Nature Singapore Pte Ltd, Singapore, 2019, 2018, pp. 123–33.
45. K. W. X. Soo, *The Role of Granularity in Causal Learning*, University of Pittsburgh, 2019.
46. V. Talanquer, Progressions in reasoning about structure – property relationships, *Chem. Educ. Res. Pract.*, 2018, **19**(4), 998–1009.
47. M. H. W. van Mil, D. Jan, B. Arend and J. Waarlo, Modelling Molecular Mechanisms : A Framework of Scientific Reasoning to Construct Molecular-Level Explanations for Cellular Behaviour, *Sci. Educ.*, 2013, **22**(1), 93–118.
48. K. M. Southard, M. R. Espindola, S. D. Zaepfel and S. Molly, Generative mechanistic explanation building in undergraduate molecular and cellular biology, *Int. J. Sci. Educ.*, 2017, **39**(13), 1795–1829.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

49. P. L. Luisi, Emergence in Chemistry: Chemistry as the Embodiment of Emergence, *Found. Chem.*, 2002, **4**(3), 183–200.
50. L. Darden, Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining, *Philos. Sci.*, 2002, **69**(S3), 354–365.
51. P. G. Mahaffy, A. Krief, H. Hopf, G. Mehta and S. A. Matlin, Reorienting chemistry education through systems thinking, *Nat. Rev. Chem.*, 2018, **2**(4), 126.
52. A. H. Johnstone, Why is science difficult to learn? Things are seldom what they seem, *J. Comput. Assist. Learn.*, 1991, **7**, 75–83.
53. P. Machamer, L. Darden and C. F. Craver, Thinking about Mechanisms, *Philos. Sci.*, 2000 Mar, **67**(1), 1–25.
54. J. Biggs and C. Tang, *Teaching for Quality Learning at University*, The Society for Research into Higher Education, 4th edn, 2011.
55. J. Biggs, Enhancing teaching through constructive alignment, *High Educ.*, 1996, **32**, 347–364.
56. V. R. Ralph, L. J. Scharlott, C. E. Schwarz, N. M. Becker and R. L. Stowe, *Beyond Instructional Practices: Characterizing Learning Environments that Support Students in Explaining Chemical Phenomena*, 2022, pp. 1–35.
57. J. Biggs and C. Tang. Aligning assessment tasks with intended learning outcomes: principles, in *Teaching for Quality Learning at University*, 2011, pp. 191–223.
58. G. Wiggins and J. McTighe, What is backward design, *Underst. Des.*, 1998, **1**, 7–19.
59. E. O'Connor, K. Roy, E. Walsh, D. Huang, D. Y. J. Ke, E. Campbell Brown, *et al.*, *The Evaluation of an Integrated Growth & Goals Module to Better Equip Students with Learning Skills in Postsecondary Courses: Systematic, Scalable, and Explicit*, June 2021.
60. M. S. Carle and A. B. Flynn, Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chem. Educ. Res. Pract.*, 2020, **21**(2), 622–637.
61. D. Wood, J. S. Bruner and G. Ross, The Role of Tutoring in Problem Solving, *J. Child Psychol. Psychiatry*, 1976, **17**(2), 89–100.
62. N. Graulich and I. Caspari, Designing a scaffold for mechanistic reasoning in organic chemistry, *Chem. Teach. Int.*, 2021, **3**(1), 19–30.
63. M. Rodemer, J. Eckhard, N. Graulich and S. Bernholt, Connecting explanations to representations: benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry, *Int. J. Sci. Educ.*, 2021, **43**(17), 2707–2728.
64. J. M. Deng and A. B. Flynn, Reasoning in scientific arguments [Internet], 2020 [cited 30 September 2021]. Available from: https://www.youtube.com/watch?v=4Gfg_BrHYLc.
65. J. M. Deng and A. B. Flynn. Scientific arguments: claim, evidence, reasoning [Internet], 2020 [cited 30 September 2021]. Available from: https://www.youtube.com/watch?v=-xOWE7Zzph8.
66. Flynn Research Group, *Scientific Argumentation*, 2021.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

67. M. K. Orgill, S. York and J. Mackellar, Introduction to Systems Thinking for the Chemistry Education Community, *J. Chem. Educ.*, 2019, **96**(12), 2720–2729.
68. A. B. Flynn, M. Orgill, F. Ho, S. York, S. A. Matlin and D. J. C. Constable, *et al.*, Future Directions for Systems Thinking in Chemistry Education: Putting the Pieces Together, *J. Chem. Educ.*, 2019, **96**(12), 3000–3005.
69. *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*, ed. S. D. Blum, West Virginia, U.S., West Virginia University Press, 2020.

Downloaded from http://books.rsc.org/books/edited-volume/chapter-pdf/1501664/bk9781839164910-00074.pdf by University of Ottawa user on 18 April 2024

Chapter 5: The role of language in students' justifications of chemical phenomena

5.1 Introduction to the research article presented in this chapter

In the research described in Chapters 2–4, we learned how students construct scientific arguments in organic chemistry, including implications for framing and communicating expectations related to argumentation in chemistry.

As we analyzed the student responses in the research described in Chapters 2 and 3, a new question began to emerge in our research group: "How many of these responses were written by students who are learning English?" All of the responses we analyzed were written in the English language, but many of students who produced these responses may be learning English-as-an-additional language (Eng+). Linguistic diversity is especially pertinent at the University of Ottawa: a bilingual institution that offers students from over 150 countries with opportunities to complete their studies in either French or English. How might students' language backgrounds be affecting their ability to communicate their thinking and construct arguments in chemistry?

If scientific argumentation is an essential skill for *all* citizens in an increasingly global society, then we need to ensure that *all* students have equitable opportunities to demonstrate and develop this skill. Without a better understanding of how students' language backgrounds played a role in how they engaged in a complex linguistic practice like scientific argumentation, we are left wondering about how equitable these types of assessment are. By learning about how equitable these types of assessments are, we can then investigate what might be changed in these types of assessments to improve equity. Therefore, Chapter 5 describes our investigation of how students' English language proficiency and history was associated with the reasoning demonstrated in their written chemistry arguments.

The research questions for this project emerged in summer 2020—around the same time, George Floyd was murdered in the United States, which led to a global anti-racism movement

that prompted many of us to reflect on the roles that we play in perpetuating inequity and exclusion in our society. COVID-19 had emerged earlier in the same year, and I was personally and deeply affected by the anti-Asian hate events that occurred during the pandemic. These two events prompted many in chemistry education research to consider how we might better incorporate equity, diversity, and inclusion (EDI) into our research and teaching. On a personal level, as a second-generation Chinese Canadian, these two events prompted me to reflect on how racism and xenophobia has impacted my life and the lives of those around me, the most notable being my parents: immigrants from China whose perspectives and expertise were constantly underappreciated because of their challenges with English.

Thank you to Malek Rahmani, an undergraduate honours student in our group at the time, who helped with the inter-rater reliability and analyzed portions of participants' language data.

5.2 Copyright

Reproduced from "Deng, J. M.; Rahmani, M.; Flynn, A. B. The Role of Language in Students' Justifications of Scientific Phenomena. *Int. J. Sci. Educ.* **2022**, *44* (13), 2131–2151. doi.org/10.1080/09500693.2022.2114299." with permission from Taylor & Francis.



International Journal of Science Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tsed20

The role of language in students' justifications of chemical phenomena

Jacky M. Deng, Malek Rahmani & Alison B. Flynn

To cite this article: Jacky M. Deng, Malek Rahmani & Alison B. Flynn (2022) The role of language in students' justifications of chemical phenomena, International Journal of Science Education, 44:13, 2131-2151, DOI: 10.1080/09500693.2022.2114299

To link to this article: https://doi.org/10.1080/09500693.2022.2114299

View supplementary material

Published online: 25 Sep 2022.

Submit your article to this journal

Article views: 496

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tsed20

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
2022, VOL. 44, NO. 13, 2131–2151
https://doi.org/10.1080/09500693.2022.2114299

Routledge
Taylor & Francis Group

Check for updates

The role of language in students' justifications of chemical phenomena

Jacky M. Deng, Malek Rahmani and Alison B. Flynn

Department of Chemistry and Biomolecular Sciences, Faculty of Science, University of Ottawa, Ottawa, Canada

ABSTRACT

Making decisions, reasoning, and constructing arguments with scientific evidence are essential skills for all members of society, especially in a world facing complex socioscientific issues such as climate change and pandemics. Argumentation is a complex linguistic practice but little is known about how students from diverse language backgrounds engage in argumentation. The goal of this study was to identify how students' English language proficiency and history was associated with the reasoning demonstrated in their written arguments. We found that students with lower English proficiency and less English history produced fewer causal responses compared to students with higher English language proficiency and history. Follow-up interviews with fifteen participants revealed that students' comfort communicating in English on assessments was affected by a combination of general and academic language experiences. Findings suggest a need to identify the barriers encountered by students from diverse language backgrounds during argumentation to ensure students from all language backgrounds have equitable supports and opportunities to demonstrate their scientific abilities.

ARTICLE HISTORY
Received 18 January 2022
Accepted 15 August 2022

KEYWORDS
Science; technology; society; reasoning; English as a second/additional language

Introduction

In a world facing complex socioscientific issues (e.g. climate change, viral pandemics, misinformation), citizens need to be able to build arguments and reason from scientific evidence (National Research Council, 2013; Organisation for Economic Cooperation and Development, 2006; Social Sciences and Humanities Research Council, 2018; United Nations, 2015). Students can develop scientific argumentation skills through classroom activities and assessments (e.g. asking students to justify why a phenomenon occurs) but the linguistic skills required to express clear arguments can be a challenge for students (McNeill et al., 2006; McNeill & Berland, 2010; McNeill & Krajcik, 2008b, 2008a). Constructing arguments in science can be especially challenging for students learning English-as-an-additional language (Eng+) who are expected to learn, use, and communicate abstract scientific concepts in academic language while they are

CONTACT Alison B. Flynn alison.flynn@uottawa.ca Department of Chemistry and Biomolecular Sciences, Faculty of Science, University of Ottawa, Ottawa, Ontario, Canada
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09500693.2022.2114299.
© 2022 Informa UK Limited, trading as Taylor & Francis Group

simultaneously learning English (Abedi et al., 2020; Afitska & Heaton, 2019; Buxton et al., 2014a; Curtis & Millar, 1988). Students' challenges with language may also be exacerbated in disciplines like chemistry in which terms like 'stability' and 'resonate' have different meanings from their everyday definitions (Childs et al., 2015; Uhl Chamot, 1995).

Eng+ students face unique linguistic challenges with *knowing*, *doing*, and *talking* science (Lee & Fradd, 1998). *Knowing* science means making meaning of scientific knowledge by connecting new information to previous knowledge. Language is foundational to learning, as it mediates how learners engage with scientific concepts and construct meaning. Learning new information can be a challenge for Eng+ students, as they may have learned prerequisite concepts in a language or culture different from English, making it more challenging to identify and establish coherent connections between new information and their previous knowledge. *Doing* science involves scientific practices, such as engaging in inquiry, proposing arguments and explanations, and interpreting and identifying evidence to make sense of the world (Lee et al., 2013; Lee & Fradd, 1998; National Research Council, 2013). Many of these practices require complex language functions and academic traditions that require sufficient levels of language proficiency within a given discipline. *Talking* science means communicating in the language of science: academic English (Lemke, 1990). Learning and communicating in academic English poses challenges for many learners, including Eng+ students who may have less experience with the expected communication patterns in science. For example, in contexts like exams where students may be asked to articulate an explanation or argument for why a phenomenon occurs, Eng+ students may engage in additional cognitive operations compared to English-as-a-first language (Eng1st) students in order to interpret English questions and then generate a written response in English (Abedi, 2002, 2015; Abedi et al., 2004; Abedi & Lord, 2001; Buxton et al., 2014a; González-Howard & McNeill, 2016; Lee, 2018; Lee & Orgill, 2021; Solano-Flores & Trumbull, 2003) (Figure 1).

The abovementioned linguistic challenges faced by Eng+ students can manifest into inequities in how Eng+ students are assessed and evaluated in science classrooms and educational research. Eng+ students' achievement in Science, Technology, Engineering, and Mathematics (STEM) fields has been found to be associated with language ability, regardless of students' abilities to grasp and apply scientific concepts (Curtis & Millar, 1988; Maerten-Rivera et al., 2010; Solano-Flores & Trumbull, 2003). Eng+ students' written responses can score lower on evaluations than those of Eng1st students due to evaluators' implicit bias (Huang, 2008; Lindsey & Crusan, 2011; Milnes & Cheng, 2008) and students' challenges communicating their knowledge (Lyon et al., 2012; Swanson et al., 2014; Wolf et al., 2008). As Kieffer et al. (2009) writes: 'Because language plays an integral role in most, if not all, academic learning, any test of academic achievement is also, to some degree, a test of language ability' (Kieffer et al., 2009).

Eng+ students form a prominent and increasing population of students in science classrooms around the world but have been historically overlooked in postsecondary science courses (Kanno & Cromley, 2013). There are over 1,250,000 international students studying in North America, the majority of whom come from nations where English is not a primary language (Project Atlas, 2019). In the United States (US), 4.6

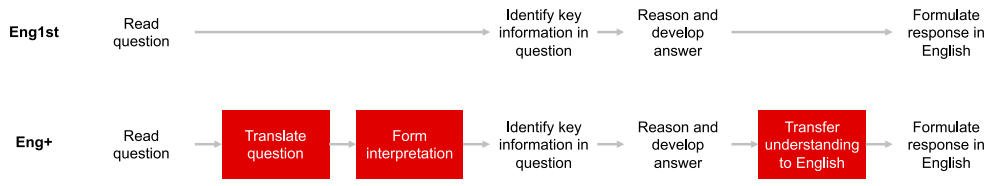


Figure 1. An outline of processes students use to solve problems, with additional processes that Eng+ students may use highlighted in red. Adapted from Lee (2018).

million Eng+ students make up about 9.4% of all classrooms, with 1.1 million Eng+ students in US postsecondary institutions (National Center for Education Statistics, 2017). Of STEM graduate students in the US, 38% are temporary visa holders from non-English nations, a number that has steadily increased since 2000 (National Science Foundation, 2020). In the United Kingdom, the Eng+ student population has increased 3.7% since 2012 (21.2% of all students) (Oxley & de Cat, 2019). In Canada, a bilingual French/English nation, the number of international students has risen from 40,000 in the late 1990s to over 340,000 in 2019, making up 16% of all university and college enrolments (Higher Education Strategy Associates, 2020). In certain Canadian institutions, students also have the option to complete their studies in French (Les CÉGEPs du Quebec, 2022; Université de Montréal, 2022; University of Ottawa, 2022a).

Neglecting the issues facing Eng+ students can result in increased attrition of students from a diversity of linguistic, cultural, and ethnic and backgrounds, and curbed development of key skills, such as scientific argumentation (Abedi, 2002; Afitska & Heaton, 2019; Allensworth, 2005; Buxton et al., 2014a, 2015; Chin, 2010; Maerten-Rivera et al., 2010; McNeil, 2005; National Academies of Science Engineering and Medicine, 2018; Pyburn et al., 2013; Solano-Flores & Trumbull, 2003; Sotelo-Dynega et al., 2013; Wolf et al., 2008; Yore et al., 2004). Such consequences may then limit the benefit that scientific and other professional communities might gain from individuals from Eng+ backgrounds (Amano et al., 2016; Cheng et al., 2019; Elnathan, 2021; Ramírez-Castañeda, 2020).

In this study, we investigated how English language proficiency and history was associated with the reasoning in students' written chemistry arguments. Much of the current work on language and learning has been conducted in K–12 mathematics, biology, or general science contexts, with few investigations in chemistry (Pyburn et al., 2013) and post-secondary contexts (LaCosse et al., 2020), despite chemistry's dependence on Anglophonic nomenclature, models, and symbols (Childs et al., 2015). Educational researchers have studied students' scientific reasoning and argumentation skills in various contexts and disciplines (Becker et al., 2016; Bodé et al., 2019; Cooper et al., 2016; Deng & Flynn, 2021; Moon et al., 2016, 2019; Osborne & Patterson, 2011; Russ et al., 2008; Watts et al., 2020; Weinrich & Talanquer, 2016); however, much of this work has relied on analyses of arguments constructed in a single language (usually English), without considering the role of language in these types of investigations.

Goals and research questions

Our research questions were:

(1) What differences in reasoning might there be in arguments constructed by postsecondary science students who have learned English-as-an-additional language (Eng+) and English-as-a-first language (Eng1st)?
(2) In what ways might English language proficiency and history be correlated with the reasoning postsecondary science students demonstrate when constructing an argument in chemistry?

In this study, we use a modes of reasoning framework to characterise the reasoning demonstrated in students' written arguments (Sevian & Talanquer, 2014). Building on previous frameworks designed to identify components of scientific argumentation (Brown et al., 2010), the modes of reasoning framework aims to characterise different levels of complexity in students' reasoning and reflect core practices of chemistry (i.e. chemical thinking) (Sevian & Talanquer, 2014). The framework has been used to characterise students' reasoning in terms of four distinct modes (descriptive, relational, linear causal, multi-component causal) that progress from listing features and/or the properties of entities to elaborating on cause-and-effect relationships between relevant properties of entities and their activities (Bodé et al., 2019; Carle et al., 2020; Caspari et al., 2018; Deng & Flynn, 2021; Moon et al., 2016; Moreira et al., 2019; Sevian & Talanquer, 2014; Toulmin, 1958; Weinrich & Talanquer, 2016).

Methods and materials

Context, recruitment, and sample

The study was conducted at the University of Ottawa, a bilingual, research-intensive institution in which students have the option to complete coursework in either English or French. All students from non-English backgrounds must complete an English-language test (e.g. TOEFL) to enrol in courses taught in English at the University; 30% of students at the University report using French as their primary language during their studies (University of Ottawa, 2020). To our knowledge, this is the first language-focused study of post-secondary students who may be completing their studies in English, French, or both. The University also enrols over 6,500 international undergraduate students (17% of the total undergraduate student population) from over 150 countries (University of Ottawa, 2020). Based on data collected using this study's instruments, 31% of participants reported majority or equal preference for a non-English language compared to English ($n = 51$). Only 26% of participants reported a complete preference for English ($n = 44$). Non-English first or preferred languages known amongst participants included French, Arabic, Chinese (Mandarin), German, Spanish, Mauritian Creole, Tagalog, Dari, Farsi, Hindi, Kinyarwanda, and Portuguese.

In their first year, students enrolled in the Faculty of Science take large (200–400 students per section) introductory courses in biology, chemistry, mathematics, *etc*. (University of Ottawa, 2022b). Following this first year, students' course sequences begin to diverge depending on their degree majors (e.g. chemistry, biochemistry, biology), with opportunities to generate written arguments on discipline-specific phenomena (e.g. 'justify your choice' questions) varying between courses and programmes. Students can complete any or all courses in either French or English.

We asked professors teaching chemistry courses at the University to forward a recruitment text to students in their courses via email. The final sample was composed of 166 participants who were students in chemistry courses across all four academic years.

Instruments and data collection procedure

In an online format, participants were asked to construct an argument in response to a social media post related to ocean acidification using chemical equilibria data (Figure 2), which was based on a prompt used in a previous study on students' argumentation in chemistry (Moon et al., 2019). The prompt was also chosen because its principal chemistry content (equilibria) is an introductory chemistry concept introduced in General Chemistry at the University of Ottawa. Participants were also asked to report demographic information, including whether they identified as Eng+ or Eng1st, the primary language in which they had learned chemistry, and their English language proficiency and history using the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007). The instruments and evidence for validity and reliability for measures can be found in the Supplementary Materials (SM).

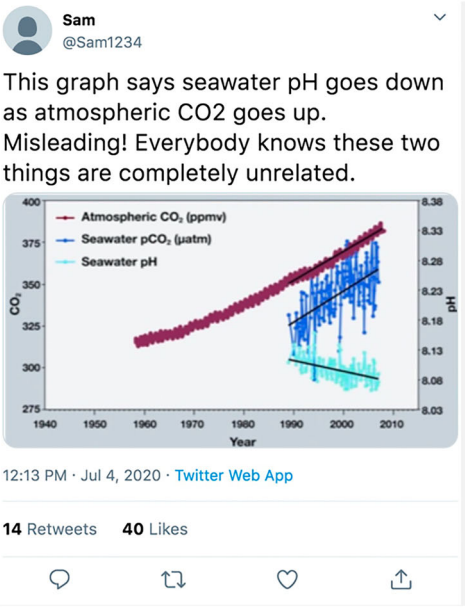
We also conducted standardised, open-ended interviews (protocol in SM) with a subset of participants ($n = 15$) to gain additional insight into participants' perceptions of how language contributes to their experiences when communicating in science and chemistry education.

Data analysis

This study followed a mixed-methods approach (Cresswell, 2012). We qualitatively analysed students' written arguments about a chemical phenomenon and employed quantitative analyses to determine whether differences in reasoning were associated with students' language backgrounds and proficiencies.

We analysed participants' arguments to identify the claim, evidence used, and mode of reasoning: descriptive, relational, linear causal, or multi-component causal (Table 1) (Bodé et al., 2019; Carle et al., 2020; Caspari et al., 2018; Deng & Flynn, 2021; Moon et al., 2016; Moreira et al., 2019; Sevian & Talanquer, 2014; Toulmin, 1958; Weinrich & Talanquer, 2016). We used this analytical framework as it requires demonstrating increasingly sophisticated linguistic complexity when progressing from descriptive to causal modes of reasoning. *Descriptive* arguments list or give features and/or the properties of entities (e.g. reactants, products) without establishing connections. *Relational* arguments include connections between properties of the entities and their activities, but these relationships are discussed in a correlative fashion (i.e. lack causality). *Causal* arguments – linear and multi-component – include all features of a relational argument and additionally contain cause-and-effect relationships between the relevant properties of the entities and their activities.

An initial coding protocol was developed by author JMD using the entire dataset ($N = 166$). Inter-rater reliability (IRR) was measured using Krippendorf's α to mitigate inflation of agreement due to chance agreements between raters (Krippendorff, 1970). The IRR process began with co-authors JMD and MR independently coding 15% ($n = 26$) of the full data set with the initial codebook. Krippendorff's α was calculated to be



$$CO_{2(aq)} + H_2O_{(l)} \rightleftharpoons H_2CO_{3(aq)}$$

$$H_2CO_{3(aq)} + H_2O_{(l)} \rightleftharpoons HCO^-_{3(aq)} + H_3O^+_{(aq)}$$

$$HCO^-_{3(aq)} + H_2O_{(l)} \rightleftharpoons CO^{2-}_{3(aq)} + H_3O^+_{(aq)}$$

$$pK_a\ H_2O_{(l)} = 14.0$$
$$pK_a\ H_2CO_{3(aq)} = 6.35$$
$$pK_a\ HCO^-_{3(aq)} = 10.3$$

Do you agree/disagree with the post? Justify your choice in an argument.

Figure 2. Participants were asked to construct an argument to the social media post (left) using relevant chemical equilibria (right).

0.372. After discussing disagreements, authors JMD and MR modified the codebook and conducted a second round of coding by coding another 15% of the full data set with the modified codebook. After this second round, Krippendorff's α was calculated to be 0.908. Throughout the process, participants' arguments were not organised in terms of Eng+ and Eng1st groups to limit bias, ensuring that coders were blind to which arguments were constructed by Eng+ and Eng1st students. No arguments demonstrated multi-component causal reasoning in the dataset and so that mode of reasoning does not appear in the codebook or subsequent description of findings.

English language profiles were created for each participant by using respondents' LEAP-Q data to develop English language proficiency and history indices (Figure S1) (Krizman et al., 2014; Marian et al., 2014; Reichle & Birdsong, 2014). Statistical tests were used to analyze how participants' language backgrounds were associated with their demonstrated mode of reasoning. All statistical tests were conducted in SPSS (details in the SM). Fischer's exact test was used to test independence of the modes of reasoning between Eng+ and Eng1st responses, followed by post-hoc pairwise *Z*-tests with Bonferroni adjusted *p*-values for multiple comparisons. Cluster analysis was conducted based on two clustering variables: relative English language history and relative English language proficiency. After evaluating parameters for collinearity (Table S2), full factorial multinomial logistic regression was used to evaluate the association between English language proficiency and history (independent covariates) and mode of reasoning (dependent categorical output). Academic year was included as a control variable (categorical predictor). Findings from the first regression analysis prompted a second logistic regression, with mode of reasoning as the dependent variable and (a) time in English language country, (b) time in English language family, and (c) time in

English language school/workplace as independent predictors. Lastly, the relationships between language group (Eng+, Eng1st), language of chemistry instruction (English, French), and reasoning demonstrated in participants' arguments (descriptive, relational, causal) were investigated.

Interviews were transcribed and analysed thematically (Cresswell, 2012; Lincoln & Guba, 1985), with attention to: (a) how Eng+ participants perceived the role of language in their own learning experiences and (b) how Eng+ and Eng1st participants perceived the role of language in others' learning experiences (peers, friends, classmates).

Results

Eng+ participants constructed fewer causal arguments than Eng+ participants

Eng+ participants constructed fewer causal arguments than Eng1st participants and instead gave more descriptive and relational arguments, χ^2 (1, N = 166) = 11.09, $p < 0.01$, $\phi = 0.253$ (small–medium effect size) (Figure 3); in this analysis, Eng1st and Eng + participants were treated as dichotomous groups. Eng+ arguments commonly described *what* was occurring (descriptive) and/or *how* the data presented to them were connected (relational). For example, Participant 28 acknowledged how the creation of H_3O^+ could be related to decreasing pH, but did not further elaborate on how or why these ideas are related:

> I disagree with Sam. The equations show clearly a relationship between the presence of CO_2 and the formation of H_3O^+. H_3O+ makes the pH of the water to go down.

In contrast, Eng1st arguments more frequently described *why* the data presented were connected and relevant to each other, often citing Le Chatelier's principle to support their argument (linear causal). For example, Participant 38:

> I disagree with Sam. According to Le Chatelier's principle as CO_2 in the atmosphere increases, CO_2 in the ocean must also increase. This means CO_2 in the first line of the series of equilibria will be pushed towards the carbonic acid product. This in turn will

Table 1. Modes of reasoning used to characterise the reasoning demonstrated in participants' arguments.

Mode of reasoning	Definitions, adapted to the prompt	Examples
Descriptive	Argument describes explicit features of the graph or equations (e.g. as atmospheric CO_2 concentration goes up, pH goes down).	Based on the equations, H_3O^+ is generated. Based on the graph, CO_2 goes up as pH goes down.
Relational	Argument identifies that increased CO_2 concentrations (atmospheric or aqueous) are correlated with H^+ being generated, using the equations. Identifies that increased H^+ is correlated with decreasing seawater pH.	Based on the graph, CO_2 goes up as pH goes down. As is shown in the equations, increasing CO_2 produces more H_3O^+ ions. Based on the equations, H_3O^+ is generated. Increasing H_3O^+ makes the seawater more acidic, so pH goes down.
Linear Causal	Argument uses Le Chatelier's principal or pK_a values to justify why atmospheric CO_2 concentrations influence the directions of the chemical equilibria and connects back to the impact that has on seawater pH.	Based on the graph, CO_2 goes up as pH goes down. Due to Le Chatelier's principle, increasing CO_2 produces carbonic acid, which leads to increased concentration of H_3O^+ ions. This results in increased acidification of the seawater, and thus a lower pH.

interact with the water producing acidic hydronium ions/hydrogen ions. The conjugate base produced from the second line of the chemical equilibria can still lose a hydrogen to a surrounding water molecule, creating more acid species. The pK_a values indicate the acid strength (how much they dissociate in water), and both of the carbonic acid species are more acidic than water, contributing to the acidification of the ocean.

Participants with lower English history and proficiency scores produced the fewest causal arguments

We encountered a challenge analyzing the data from the Eng+ group because participants could have a range of histories and proficiencies, including: if they were relatively new to English, experienced with English-as-an-additional language, and if they considered English as their more dominant language even if they learned it as an additional language. Heterogeneity in the Eng+ and Eng1st samples (identified in the LEAP-Q data) is a common challenge when categorising individuals into language groups (Lee & Stephens, 2020; National Academies of Science Engineering and Medicine, 2018). To better capture the heterogeneity between groups, we used cluster analysis using the LEAP-Q data. This analysis uncovered four language clusters: (1) high English history and proficiency; (2) high English history and low proficiency; (3) low English history and high proficiency; and (4) low English history and proficiency (Figure 4).

Cluster 1 responses more frequently exhibited causal reasoning compared to Cluster 4, χ^2 (6, N = 166) = 14.02, $p < 0.05$, $\phi = 0.284$ (large effect size) (Figure 5a). The proportion of causal reasoning decreased from Cluster 1 to Cluster 4, while the proportion of descriptive and relational reasoning increased.

Cluster 1 had the highest proportion of participants who identified as Eng1st (99%), χ^2 (9, N = 166) = 94.97, $p < 0.0001$, $\phi = 0.756$ (large effect size) (Figure 5b). Within Cluster 1, 91% had learned chemistry primarily in English (EngCHM) and 8% learned chemistry primarily in French (FrCHM). In contrast, Cluster 4 had the highest proportion of Eng+ participants (79%), with 42% having learned chemistry in English and 37% in French.

Higher English history was predictive of generating a causal argument

English language history was a significant predictor of exhibiting causal reasoning relative to the reference category (relational reasoning), controlling for academic year, Exp (B) = 1.057, $p < 0.05$ (Table S3). That is, participants who had more experiences and time with English were more likely to produce a causal response compared to a relational response. More time in an English language family was a significant predictor of exhibiting causal reasoning compared to a relational response, regardless of academic year, Exp (B) = 1.085, $p < 0.05$ (Table S4).

Eng+ participants who had learned chemistry primarily in French did not demonstrate causal reasoning at all in their arguments (Figure 6). The proportion of causal arguments increased as the alignment between language group and language of instruction increased: Eng+ participants in English chemistry (17% causal), Eng1st participants in French chemistry (37% causal), and Eng1st participants in English chemistry (44% causal).

Preferred language for communicating in chemistry was connected to general and chemistry-specific language experiences

Most participants preferred communicating chemistry in the language of chemistry instruction, regardless of their first language (Figure 7, with examples in Table S8). Both Eng1st and Eng+ participants who had learned chemistry primarily in English ($n = 5$ and $n = 4$, respectively) said they preferred English in chemistry contexts. They said it would be difficult for them to communicate the same ideas in their first and/or home languages. For example, Steph, a student who spoke Urdu and Punjabi as their first language at home, when asked how they would communicate chemical ideas in these languages:

> I would definitely use a lot of English words, just because I wouldn't know the words for the things I want to talk about in Urdu or Punjabi. So, you would definitely hear a lot of English … Like, the Urdu and Punjabi I know is more just conversational. Not really scientific.

Two participants who had learned chemistry in English but spoke French in social and home contexts said they could also communicate chemistry effectively in French given the bilingual context ($n = 2$). These participants also had high English history and proficiency scores. They said they thought about chemistry in English and could translate this knowledge into French if needed due to prior experiences learning some chemistry in French. For example, Jen stated:

> I did my schooling all in French, all the way up to second year. And then like, I just went full English. My dad and his whole side of the family is English, so I'm pretty equally comfortable with both languages … I find that in courses like biology and stuff, my brain still functions in French because I took those courses in French. When I'm doing chemistry, my brain thinks in English, so it actually comes easier to me in English, because I have more experience … in French I think it might even be a bit harder, because I'd have to translate some of the words, but I mean, I wouldn't say that it'd be hard. I could still do it easily. But I would have to mentally translate.

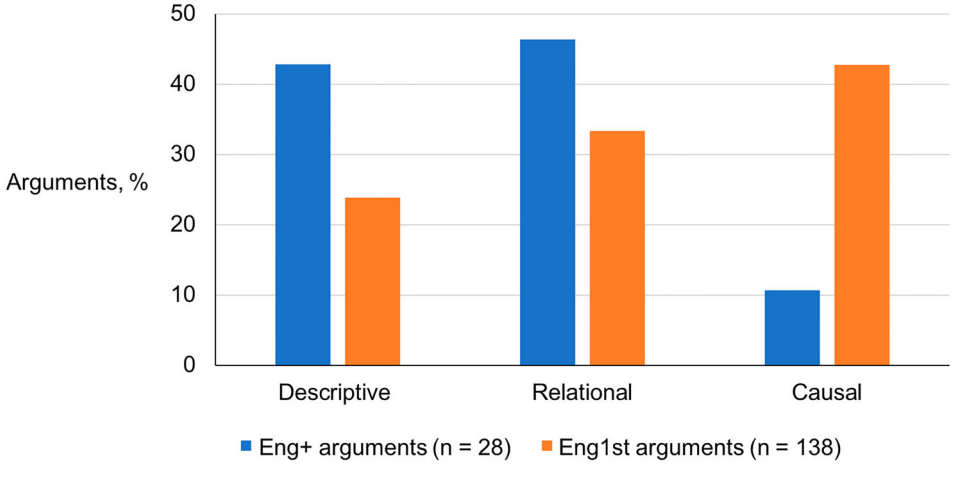


Figure 3. Reasoning exhibited in scientific arguments. Eng+ ($n = 28$, blue, left) and Eng1st ($n = 138$, orange, right).

Eng1st participants learning chemistry in French said they often thought about chemistry in French and could more easily generate a response in French compared to English (n = 3). They believed French allowed them to communicate more effectively and provide more detail than English. Melanie stated:

> I've done my whole schooling in French, so all the terms and stuff are in French in my head … In French, instead of saying, 'I would disagree with the Twitter user, it is not misleading', I would explain why it was not misleading, instead of just saying it. Because in English, I feel like I affirm more things instead of explaining why it's not as misleading. So, I would have probably went into the why a little bit more.

These findings were reinforced by interview participants' LEAP-Q data. Participants who exhibited relatively higher levels of English history and proficiency stated that they were confident in their abilities to communicate chemistry in English (n = 5). In contrast, describing a need to mentally translate chemistry knowledge from French to English or vice versa was associated with lower English history (n = 5), while a preference for French over English was associated with relatively lower scores of both English history and proficiency (n = 3).

The interview findings were reinforced when Eng+ respondents were given the opportunity to respond given the opportunity to respond to the ocean acidification prompt in their first language (n = 30). Eng+ participants from French-first backgrounds with experience using French in chemistry contexts felt that they could either argue better in French (n = 8) or at a similar level in both French and English (n = 8). Two students shared:

> I liked having the option to answer in my native language! I've done Organic I and II in both French and English so I think I could manage if I weren't given a choice, but allowing me to answer in whichever language I am most comfortable in makes me feel more confident in my response and abilities.

> [My response] would be the same as usual because uOttawa is French/English bilingual like me.

In contrast, Eng+ participants who had limited experience using their first language in chemistry contexts preferred English (n = 10), for example:

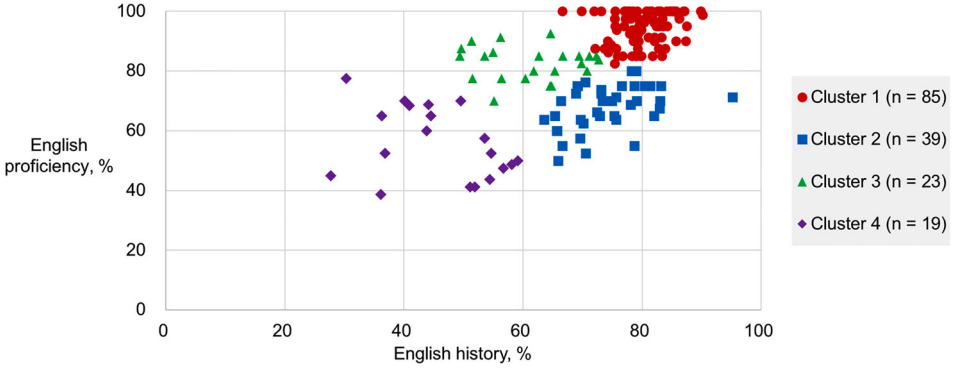


Figure 4. Participants' English proficiency and history, organised into four clusters using cluster analysis (N = 166).

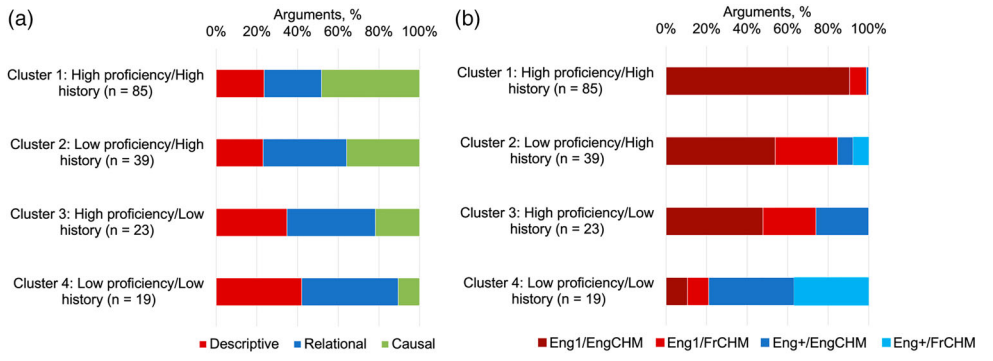


Figure 5. (a) Distributions of the modes of reasoning for each language cluster; (b) Distributions of general/chemistry language backgrounds for each language cluster.

> I wouldn't be able to [respond in Spanish] as I have done all of my higher educationin English and have therefore not learned the proper terminology to have an educated conversation in Spanish about this topic.
>
> To be honest, despite French being my first language, I found it very hard to come up with the proper words to use because I've become accustomed to discussing Chemistry in English.

Discussion

Supporting students' reasoning abilities requires supporting both ability and expression

Three main findings emerge from this study: (1) Eng+ participants produced fewer arguments with causal reasoning compared to Eng1st participants; (2) participants with lower English language proficiency and history produced fewer causal responses compared to participants with higher English language proficiency and history; and (3) participants often preferred communicating chemistry in the same language as they had learned it.

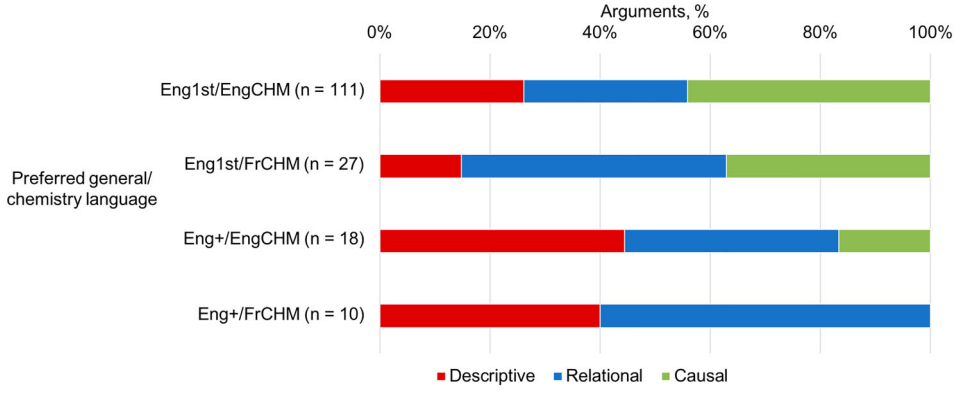


Figure 6. Intersection between language group (Eng+, Eng1st), primary language of chemistry instruction (EngCHM, FrCHM) and reasoning.

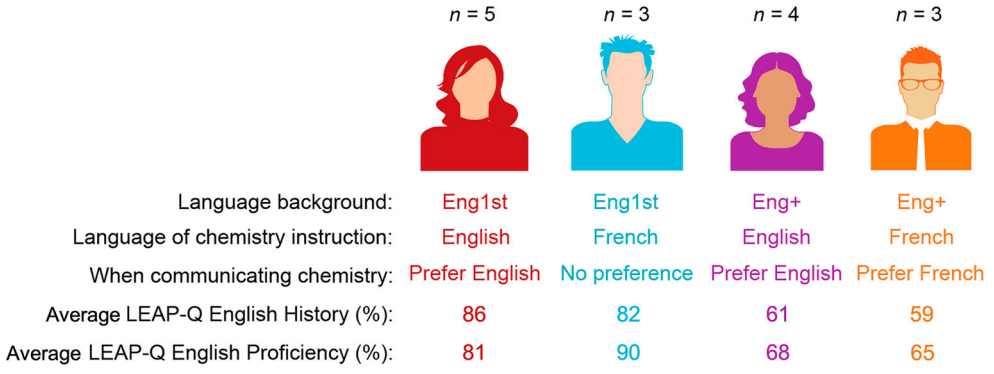


Figure 7. Based on interview data, participants' comfort with responding in English vs. French was connected to both their general language background (Eng1st or Eng+) and their primary language of academic chemistry instruction (English or French).

Taken together, the findings suggest that language ability is correlated with the reasoning demonstrated in students' written assessments in chemistry, reinforcing previous findings about Eng+ students' challenges generating written arguments and explanations of phenomena (Abedi, 2002, 2015; Abedi et al., 2004; Abedi & Lord, 2001; Buxton et al., 2014a; González-Howard & McNeill, 2016; Lee, 2018; Lee & Orgill, 2021; Solano-Flores & Trumbull, 2003).

However, our findings reveal that *students who are less proficient in English may still prefer to use English* for science communication. Students who are less proficient in English and/or consider themselves Eng+ and speak a non-English language at home may still prefer to communicate academic ideas in English based on their experiences using English in academic settings. This finding is reinforced by previous literature distinguishing between English for academic purposes and English for general/social purposes (Cummins, 1981; Flowerdew, 2019; Hyland, 2016; Lee & Fradd, 1998; Myles & Cheng, 2003). Therefore, certain linguistic academic accommodations, such as bilingual dictionaries, may not be helpful for Eng+ learners. Instead, other academic supports may be more useful to achieve equity; more research is needed.

Participants who had the fewest English experiences (both at home and in academic settings) produced the lowest proportion of causal arguments and felt less comfortable using English in an academic science setting. This finding suggests that students from majority non-English backgrounds may need the most support when expected to communicate academic knowledge in English. In many contexts, science may be taught entirely in a non-English language; for example, in Québec, Canada, students have the option to complete the entirety of their science education in French (Les CÉGEPs du Quebec, 2022). Individuals from these backgrounds may have little experience communicating in English, either in general or academic contexts, consequently limiting their participation in science or making participation much more challenging (Cheng et al., 2019; Elnathan, 2021; Ortega, 2021). Without support, scientists from Eng+ backgrounds face unique financial and career costs in having to navigate a scientific enterprise dominated by English (Cheng et al., 2019; Elnathan, 2021; Hanauer et al., 2019; Hanauer & Englander, 2011; Ortega, 2021; Ramírez-Castañeda, 2020).

Limitations

This study used convenience sampling so self-selection bias may have been present. Eng+ participants in this study may not be representative of the full diversity within the Eng+ student population, though we included statements in our recruitment texts to explicitly invite students from linguistically diverse backgrounds. Based on interview and LEAP-Q data, most of the Eng+ participants appeared to be long-term English learners and/or English/French bilingual students whose experiences may be distinct compared to English newcomers. Future endeavours involving Eng+ students could employ culturally-sensitive recruitment methods to invite participation from a diverse cohort of the Eng+ population (Liamputtong, 2008).

The low-stakes context (i.e. voluntary questionnaire) may have contributed to participants responding differently than in higher-stakes contexts (e.g. an exam). For example, Eng+ participants may have interpreted the expectations of the prompt/context differently than Eng1st participants, which may have shaped how they responded in the study. During the interviews, both Eng+ and Eng1st participants who demonstrated descriptive or relational reasoning stated that they would have either provided more detail or tailor their response to course expectations on an exam (Table S7). One Eng+ participant who produced a causal response explained how their self-perceived limitations with English often pushed them to provide greater detail and clarity on assessments to ensure they were being understood correctly. Despite this finding, we hesitate to make general claims about how Eng+ and Eng1st participants may have perceived the stakes differently given the limited sample size and lack of concrete evidence on how participants would respond in a higher-stakes context. In general, the statistical significance of observations from the interviews is limited by the small sample sizes and more research is needed to investigate the interplay of the various factors involved.

Other demographic factors may have contributed to how participants responded, including grade-point average, socioeconomic status, racial/ethnic background, immigrant/newcomer status, based on previous research (Afitska & Heaton, 2019; González-Howard & Suárez, 2021; LaCosse et al., 2020; Lee, 2005). Although we included participants' academic years in our logistic regression model, we did not account for participants' previous educational experiences with argumentation and reasoning in chemistry. Though some longitudinal studies have been conducted in chemistry education (Atkinson & Bretz, 2021; Crandell et al., 2019), more research is needed to determine how students' reasoning in chemistry may change as a result of different educational experiences over the course of their degree programmes. Nevertheless, important aspects of language were uncovered that merit further study.

Conclusions and implications

Constructing scientific arguments is an essential skill for members of society, especially in a world facing complex socioscientific issues. However, scientific argumentation is a complex linguistic practice and there is limited research on how language proficiency is related to individuals' argumentation skills. Therefore, we investigated how undergraduate science students' English language proficiency and history were associated with their reasoning in scientific arguments using a prompt on ocean acidification.

We found that participants' English language experiences, including their proficiency and history, were associated with the type of reasoning demonstrated. Fewer scientific arguments from English-as-an-additional language (Eng+) participants demonstrated causal reasoning compared to arguments from English-as-a-first language (Eng1st) participants. Participants with high English proficiency and history (Cluster 1) produced more arguments with causal reasoning than participants with low English proficiency and history (Cluster 4). Regression analyses revealed that English language history was a significant predictor of the reasoning exhibited in participants' arguments.

In interviews, participants described how their language preferences depended on the contexts in which they had learned a particular language. Participants described a preference for using their primary language of *chemistry* instruction when engaging with chemistry content, *even if* the language of instruction was not their preferred language in general contexts. Participants who had the fewest English experiences (both out of and in school) felt the least comfortable using English in an academic science setting.

Participating in science requires being able to communicate disciplinary content knowledge in academic English, and language barriers faced by Eng+ learners may hinder their abilities to demonstrate both disciplinary content knowledge and scientific practices, such as argumentation. Implications may also extend to the state of diversity in science more generally; for example, Eng+ professionals who have learned chemistry in non-English contexts may not be able to fully express themselves in English. That gap between ability and expression may not only hinder a person's career (Cheng et al., 2019; Elnathan, 2021), but also limit the benefit that the scientific (or other) community could gain from their expertise (Amano et al., 2016; Ramírez-Castañeda, 2020).

Future research may investigate these research questions in higher stakes settings (e.g. exams) and identify other variables that may be at play (e.g. previous educational experiences with argumentation). Studies could investigate specific challenges that exist, the nature and timing of barriers, the approaches Eng+ students use to succeed in English environments, and the effects of equity-intended interventions. The results of these studies could be used to better support Eng+ students through interventions (e.g. educational opportunities) or training for educators (e.g. designing inclusive assessments, mitigating implicit bias when evaluating work).

Academic accommodations to support Eng+ students have been proposed but need further study, including providing students with more time to complete assessments, incorporating scaffolds and visual aids on assessment items, and low-stakes opportunities to develop language and argumentation skills (Abedi et al., 2004, 2020; Cikmaz et al., 2021; Francis et al., 2006; Lee & Orgill, 2021; Ryoo & Bedell, 2019; Siegel, 2007). Others have suggested avenues focused on universal design (Thompson & Thurlow, 2002). Future research could also investigate impacts of strategies that frame Eng+ students' language backgrounds as assets and not deficits, viewing Eng+ students as a diverse group of learners who bring unique multilingual perspectives to the classroom (González-Howard & Suárez, 2021; Lee, 2021).

Creating a more equitable and inclusive society requires mitigating barriers and treating diversity as an asset to support participation of individuals from all cultures and experiences. By better understanding language-related barriers to learning and participating in science, we can better support students and scientists throughout their careers, enhancing both equity and innovation in science (Bell et al., 2011; Harris

et al., 2020; Hofstra et al., 2020; Impellizzeri & Coe, 2021; Lee et al., 2013; McKinsey & Company, 2015; Valantine et al., 2015).

Acknowledgements

JD thanks Canada's Social Science and Humanities Research Council (SSHRC) under the Vanier Graduate Scholarship program. The authors thank SSHRC for support through an Insight Grant.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Social Sciences and Humanities Research Council of Canada: Insight Grant. JLD thanks the Natural Sciences and Engineering Research Council of Canada for support through a Vanier Scholarship.

Ethical statement

The University of Ottawa's Office of Research Ethics and Integrity approved the research described in this article (H11-18-1363).

ORCID

Jacky M. Deng http://orcid.org/0000-0002-2633-6624
Alison B. Flynn http://orcid.org/0000-0002-9240-1287

References

Abedi, J. (2002). Assessment and accommodations of English language: Issues, concerns, and recommendations. *Journal of School Improvement*, *3*(1), 83–89.

Abedi, J. (2015). Language issues in item development. *Handbook of Test Development, Civ*, 355–373. https://doi.org/10.4324/9780203874776.ch17

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*(1), 1–28. https://doi.org/10.3102/00346543074001001

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234. https://doi.org/10.1207/S15324818AME1403

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. *Educational Measurement: Issues and Practice*, *39*(4), 41–52. https://doi.org/10.1111/emip.12328

Afitska, O., & Heaton, T. J. (2019). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, *103*(6), 1396–1422. https://doi.org/10.1002/sce.21545

Allensworth, E. (2005). Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004. In *Consortium on Chicago school research* (Issue January). papers://e97baf73-a10d-438d-985b-bee59744e11f/Paper/p939

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages are still a major barrier to global science. *PLoS Biology*, *14*(12), e2000933. https://doi.org/10.1371/journal.pbio.2000933

Atkinson, M. B., & Bretz, S. L. (2021). Measuring changes in undergraduate chemistry students' reasoning with reaction coordinate diagrams: A longitudinal, multi-institution study. *Journal of Chemical Education*, *98*(4), 1064–1076. https://doi.org/10.1021/acs.jchemed.0c01419

Becker, N., Noyes, K., & Cooper, M. M. (2016). Characterizing students' mechanistic reasoning about London dispersion forces. *Journal of Chemical Education*, *93*(10), 1713–1724. https://doi.org/10.1021/acs.jchemed.6b00298

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, *37*(3), 709–743. https://doi.org/10.1177/0149206310365001

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, *15*(3), 142–174. https://doi.org/10.1080/10627197.2010.530562

Buxton, C., Allexsaht-Snider, M., Aghasaleh, R., Kayumova, S., Kim, S.-H., Choi, Y.-J., & Cohen, A. (2014a). Potential benefits of bilingual constructed response science assessments for understanding bilingual learners' emergent use of language of scientific investigation practices. *Double Helix*, *2*(September 2016), 1–21. https://doi.org/10.37514/DBH-J.2014.2.1.05

Buxton, C. A., Salinas, A., Mahotiere, M., Lee, O., & Secada, W. (2015). Fourth-grade emergent bilingual learners' scientific reasoning complexity, controlled experiment practices, and content knowledge when discussing school, home, and play contexts. *Teachers College Record*, *117*(2).

Carle, M. S., El Issa, R., Pilote, N., & Flynn, A. B. (2020). Ten essential delocalization learning outcomes: How well are they achieved? *ChemRxiv*, 1–28. https://doi.org/10.26434/chemrxiv.13322771.v2

Caspari, I., Kranz, D., & Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chemistry Education Research and Practice*, *19*(4), 1117–1141. https://doi.org/10.1039/C8RP00131F

Cheng, Y., Dharwadkar, S., Sheridan, V., Rojas, C. R., Bosch Grau, M., & Gordin, M. (2019). Science's language barrier. *Nature*, *570*(7760), 265–267. https://doi.org/10.1038/d41586-019-01797-0

Childs, P. E., Markic, S., & Ryan, M. C. (2015). The role of language in teaching and learning chemistry. In J. Garcia-Martinez & E. Serrano-Torregrossa (Eds.), *Chemistry education: Best practices, opportunities and trends* (pp. 421–445). Wiley-VCH Verlag GmbH & Co. KGaA.

Chin, W. Y. (2010). Linguistic profiling in education: How accent bias denies equal educational opportunities to students of color. *The Scholar*, *12*(355), 356–384.

Cikmaz, A., Fulmer, G., Yaman, F., & Hand, B. (2021). Examining the interdependence in the growth of students' language and argument competencies in replicative and generative learning environments. *Journal of Research in Science Teaching*, *58*(10), 1457–1488. https://doi.org/10.1002/tea.21715

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating students' reasoning about acid–base reactions. *Journal of Chemical Education*, *93*(10), 1703–1712. https://doi.org/10.1021/acs.jchemed.6b00417

Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2019). Reasoning about reactions in organic chemistry: Starting it in general chemistry. *Journal of Chemical Education*, *96*(2), 213–226. https://doi.org/10.1021/acs.jchemed.8b00784

Cresswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of Education*, *163*(1), 16–29. https://doi.org/10.1177/002205748116300104

Curtis, S., & Millar, R. (1988). Language and conceptual understanding in science: A comparison of English and Asian language speaking children. *Research in Science & Technological Education*, *6*(1), 61–77. https://doi.org/10.1080/0263514880060106

Deng, J. M., & Flynn, A. B. (2021). Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items. *Chemistry Education Research and Practice*, *22*(3), 749–771. https://doi.org/10.1039/D0RP00320D

Elnathan, R. (2021). *English is the language of science — But precision is tough as a non-native speaker*. Nature. https://www.nature.com/articles/d41586-021-00899-y

Flowerdew, J. (2019). The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. *Language Teaching*, *52*(2), 249–260. https://doi.org/10.1017/S0261444819000041

Francis, D. J., Rivera, M., Leseaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. 2006(November 7). http://www.centeroninstruction.org/files/ELL1-Interventions.pdf

González-Howard, M., & McNeill, K. L. (2016). Learning in a community of practice: Factors impacting English-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, *53*(4), 527–553. https://doi.org/10.1002/tea.21310

González-Howard, M., & Suárez, E. (2021). Retiring the term English language learners: Moving toward linguistic justice through asset-oriented framing. *Journal of Research in Science Teaching*, *58*(5), 749–752. https://doi.org/10.1002/tea.21684

Hanauer, D. I., & Englander, K. (2011). Quantifying the burden of writing research articles in a second language: Data from Mexican scientists. *Written Communication*, *28*(4), 403–416. https://doi.org/10.1177/0741088311420056

Hanauer, D. I., Sheridan, C. L., & Englander, K. (2019). Linguistic injustice in the writing of research articles in English as a second language: Data from Taiwanese and Mexican researchers. *Written Communication*, *36*(1), 136–154. https://doi.org/10.1177/0741088318804821

Harris, R. B., Mack, M. R., Bryant, J., Theobald, E. J., & Freeman, S. (2020). Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a "hyperpersistent zone". *Science Advances*, *6*(24), 1–9. https://doi.org/10.1126/sciadv.aaz5687

Higher Education Strategy Associates. (2020). *The state of post-secondary education in Canada*.

Hofstra, B., Kulkarni, V. V., Galvez, S. M. N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments?—A generalizability theory approach. *Assessing Writing*, *13*(3), 201–218. https://doi.org/10.1016/j.asw.2008.10.002

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005

Impellizzeri, S., & Coe, I. R. (2021). The complex chemistry of diversity and inclusion: A 30-year synthesis. *Canadian Journal of Chemistry*, *99*(8), 653–660. https://doi.org/10.1139/cjc-2021-0063

Kanno, Y., & Cromley, J. G. (2013). English language learners' access to and attainment in postsecondary education. *TESOL Quarterly*, *47*(1), 89–121. https://doi.org/10.1002/tesq.49

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, *79*(3), 1168–1201. https://doi.org/10.3102/0034654309332490

Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, *30*(1), 61–70. https://doi.org/10.1177/001316447003000105

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, *128*(1), 34–40. https://doi.org/10.1038/jid.2014.371

LaCosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-belonging intervention improves STEM outcomes for students who speak English as a second language. *Science Advances*, *6*(40), 1–11. https://doi.org/10.1126/sciadv.abb6543

Lee, E. N. (2018). *Breaking the language barrier: Equitable assessment in general chemistry* [Doctoral dissertation, University of Nevada]. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2018-58618-069&lang=de&site=ehost-live

Lee, E. N., & Orgill, M. (2021). Toward equitable assessment of English language learners in general chemistry: Identifying supportive features in assessment items. *Journal of Chemical Education*, *99*(4), 35–48. https://doi.org/10.1021/acs.jchemed.1c00370

Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research*, *75*(4), 491–530. https://doi.org/10.3102/00346543075004491

Lee, O. (2021). Asset-oriented framing of science and language learning with multilingual learners. *Journal of Research in Science Teaching*, *58*(7), 1073–1079. https://doi.org/10.1002/tea.21694

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-language backgrounds. *Educational Researcher*, *27*(4), 12–21. https://doi.org/10.3102/0013189X027004012

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English language arts and mathematics. *Educational Researcher*, *42*(4), 223–233. https://doi.org/10.3102/0013189X13480524

Lee, O., & Stephens, A. (2020). English learners in STEM subjects: Contemporary views on STEM subjects and language with English learners. *Educational Researcher*, *49*(6), 426–432. https://doi.org/10.3102/0013189X20923708

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation.

Les CÉGEPs du Quebec. (2022). *What is a CÉGEP?* https://www.cegepsquebec.ca/en/cegeps/presentation/what-is-a-cegep/

Liamputtong, P. (2008). *Doing cross-cultural research: Ethical and methodological perspectives*.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Sage.

Lindsey, P., & Crusan, D. (2011). How faculty attitudes and expectations toward student nationality affect writing assessment. *Across the Disciplines*, *8*(4), 1–19. https://doi.org/10.37514/atd-j.2011.8.4.23

Lyon, E. G., Bunch, G. C., & Shaw, J. M. (2012). Navigating the language demands of an inquiry-based science performance assessment: Classroom challenges and opportunities for English learners. *Science Education*, *96*(4), 631–651. https://doi.org/10.1002/sce.21008

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education*, *94*(6), 937–962. https://doi.org/10.1002/sce.20408

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., Chabal, S., Bartolotti, J., Bradley, K., & Hernandez, A. E. (2014). Differential recruitment of executive control regions during phonological competition in monolinguals and bilinguals. *Brain and Language*, *139*, 108–117. https://doi.org/10.1016/j.bandl.2014.10.005

McKinsey & Company. (2015). *Why diversity matters*. https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters#

McNeil, L. (2005). Faking equity: High-stakes testing and the education of Latino youth. In A. Valanzeula (Ed.), *Leaving children behind: How "Texas style" accountability fails Latino youth* (pp. 57–112). University of New York Press.

McNeill, K. L., & Berland, L. K. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793. https://doi.org/10.1002/sce.20402

McNeill, K. L., & Krajcik, J. (2008a). Inquiry and scientific explanations: Helping students use evidence and reasoning. In *Science as inquiry in the secondary setting* (pp. 121–134).

McNeill, K. L., & Krajcik, J. (2008b). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, *45*(1), 53–78. https://doi.org/10.1002/tea

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

Milnes, T., & Cheng, L. (2008). Teachers' assessment of ESL students in mainstream classes: Challenges, strategies, and decision-making. *TESL Canada Journal*, *26*(1), 49. https://doi.org/10.18806/tesl.v26i1.129

Moon, A., Moeller, R., Gere, A. R., & Shultz, G. V. (2019). Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification. *Chemistry Education Research and Practice*, *20*(3), 484–494. https://doi.org/10.1039/c9rp00005d

Moon, A., Stanford, C., Cole, R., & Towns, M. H. (2016). The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry. *Chemistry Education Research and Practice*, *17*(2), 353–364. https://doi.org/10.1039/c5rp00207a

Moreira, P., Marzabal, A., & Talanquer, V. (2019). Using a mechanistic framework to characterise chemistry students' reasoning in written explanations. *Chemistry Education Research and Practice*, *20*(1), 120–131. https://doi.org/10.1039/c8rp00159f

Myles, J., & Cheng, L. (2003). The social and cultural life of non-native English speaking international graduate students at a Canadian university. *Journal of English for Academic Purposes*, *2*(3), 247–263. https://doi.org/10.1016/S1475-1585(03)00028-6

National Academies of Science Engineering and Medicine. (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. The National Academies Press. https://doi.org/10.17226/25182

National Center for Education Statistics. (2017). *The condition of education: English language learners in public schools*.

National Research Council. (2013). *Next generation science standards*. https://www.nextgenscience.org/

National Science Foundation. (2020). *2019 doctorate recipients from U.S. universities: Vol. NSF 21-308* (Issue December 2020). http://www.nsf.gov/statistics/2016/nsf16300/digest/nsf16300.pdf

Organisation for Economic Cooperation and Development. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA* 2006.

Ortega, R. P. (2021). Science's English dominance hinders diversity—But the community can work toward change. *Science*. https://www.sciencemag.org/careers/2020/10/science-s-english-dominance-hinders-diversity-community-can-work-toward-change

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438

Oxley, E., & de Cat, C. (2019). A systematic review of language and literacy interventions in children and adolescents with English as an additional language (EAL). *The Language Learning Journal*, *49*(0), 265–287. https://doi.org/10.1080/09571736.2019.1597146

Project Atlas. (2019). *Research and insights: Global mobility trends*. https://www.iie.org/Research-and-Insights/Project-Atlas/Explore-Data/Infographics/2019-Project-Atlas-Infographics

Pyburn, D. T., Pazicni, S., Benassi, V. A., & Tappin, E. E. (2013). Assessing the relation between language comprehension and performance in general chemistry. *Chemistry Education Research and Practice*, *14*(4), 524–541. https://doi.org/10.1039/c3rp00014a

Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS ONE*, *15*(9 September), e0238372. https://doi.org/10.1371/journal.pone.0238372

Reichle, R. V., & Birdsong, D. (2014). Processing focus structure in L1 and L2 French: L2 proficiency effects on ERPs. *Studies in Second Language Acquisition*, *36*(3), 535–564. https://doi.org/10.1017/S0272263113000594

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

Ryoo, K., & Bedell, K. (2019). Supporting linguistically diverse students' science learning with dynamic visualizations through discourse-rich practices. *Journal of Research in Science Teaching*, *56*(3), 270–301. https://doi.org/10.1002/tea.21519

Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: A learning progression on chemical thinking. *Chemistry Education Research and Practice*, *15*(1), 10–23. https://doi.org/10.1039/c3rp00111c

Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching*, *44*(6), 864–881. https://doi.org/10.1002/tea.20176

Social Sciences and Humanities Research Council. (2018). *Truth under fire in a post-fact world*. https://horizons.gc.ca/en/2018/10/19/the-next-generation-of-emerging-global-challenges/#truth-under-fire

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*(2), 3–13. https://doi.org/10.3102/0013189X032002003

Sotelo-Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. (2013). English language proficiency and test performance: An evaluation of bilingual students with the Woodcock-Johnson III tests of cogntitive abilities. *Psychology in the Schools*, *50*(8), 781–797. https://doi.org/10.1002/pits

Swanson, L. H., Bianchini, J. A., & Lee, J. S. (2014). Engaging in argument and communicating information: A case study of English language learners and their science teacher in an urban high school. *Journal of Research in Science Teaching*, *51*(1), 31–64. https://doi.org/10.1002/tea.21124

Thompson, S., & Thurlow, M. (2002). Universally designed assessments: Better tests for everyone! In *NCEO policy directions*.

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press. https://doi.org/10.1080/00048405985200191

Uhl Chamot, A. (1995). Implementing the cognitive academic language learning approach: CALLA in Arlington, Virginia. *Bilingual Research Journal*, *19*(3 & 4), 379–394. https://doi.org/10.1080/15235882.1995.10162680

United Nations. (2015). *Transforming our world: the 2030 agenda for sustainable development*. https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Université de Montréal. (2022). *Facts and figures*. https://www.umontreal.ca/en/udem/#facts-and-figures

University of Ottawa. (2020). *Quick facts 2020*. https://www.uottawa.ca/institutional-research-planning/resources/facts-figures/quick-facts

University of Ottawa. (2022a). *Francophonie and bilingualism*. https://www.uottawa.ca/francophonie-bilingualism/

University of Ottawa. (2022b). *Undergraduate studies in Science*. https://www2.uottawa.ca/faculty-science/programs/undergraduate

Valantine, H. A., Collins, F. S., & Verma, I. M. (2015). National institutes of health addresses the science of diversity. *Proceedings of the National Academy of Sciences*, *112*(40), 12240–12242. https://doi.org/10.1073/pnas.1515612112

Watts, F. M., Schmidt-Mccormack, J. A., Wilhelm, C. A., Karlin, A., Sattar, A., Thompson, B. C., Gere, A. R., & Shultz, G. V. (2020). What students write about when students write about mechanisms: Analysis of features present in students' written descriptions of an organic reaction mechanism. *Chemistry Education Research and Practice*, *21*(4), 1148–1172. https://doi.org/10.1039/c9rp00185a

Weinrich, M. L., & Talanquer, V. (2016). Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product. *Chemistry Education Research and Practice*, *17*(2), 394–406. https://doi.org/10.1039/c5rp00208g

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses (CRESST Report 732). *Review Literature and Arts of the Americas*, 1–96.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly*, *39*(3), 347–352. https://doi.org/10.1598/RRQ.39.3.8

Chapter 6: “I am working 24/7, but I can't translate that to you”: The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language backgrounds

6.1 Introduction to the research article presented in this chapter

In the research described in Chapter 5, we learned that students’ language backgrounds were associated with the reasoning they demonstrated in their written chemistry arguments. We also learned from qualitative interviews that students may still prefer to use English for scientific purposes, even if it is not the language they are most proficient in.

Chapter 6 details our research on the experiences of chemistry trainees learning English-as-an-additional language (Eng+). This research is imperative in a chemistry community filled with linguistic diversity and striving to be more equitable and inclusive. By better understanding the experiences of chemists from diverse language backgrounds, we can explore ways to mitigate barriers to equity and inclusion and enhance diversity and innovation in science. We conducted focus groups and interviews with 18 Eng+ chemistry trainees at the University of Ottawa, focused on three research questions: (1) What are Eng+ chemistry trainees’ language experiences before entering English-dominant research programs? (2) What language barriers have trainees faced in learning, communicating, and doing chemistry? and (3) What strategies and supports have been helpful or needed?

The research described in Chapter 6 marks a shift in the direction of my doctoral research. Though I enjoyed the projects focused on argumentation and reasoning in chemistry, I felt myself being pulled more and more in the direction of projects focused on EDI in chemistry education. I could also feel myself becoming more independent in research—not only in my skills and expertise, but also in identifying and articulating the research interests and questions I cared about. I want to explicitly express my gratitude to Dr. Flynn here. While constantly pushing me to

think about “Why is this work or research question important?”, (I think) she also recognized my growth as a researcher and made the conscious decision to give me more freedom to explore my interests at this stage in my development.

This is without a doubt the doctoral research project I am the proudest of. I have had opportunities to share this research at group meetings, conferences, and seminars across North America (see List of Contributions, pages xiv–xviii), as well as *Chemical & Engineering News (C&EN),* a prominent chemistry magazine (Vasquez, 2023). I am always humbled when researchers—in chemistry and beyond—tell me that this work has resonated with their experience and helped them feel represented and heard.

This project was also transformative on a personal level. There had been moments prior to this project where I felt like someone with my cultural experiences and perspectives—a first-generation student from a Chinese immigrant family—did not belong in chemistry education, research, or academia. This project (and Dr. Flynn’s belief in my ability to lead it) instilled within me a self-belief that my identity, experiences, and knowledge are valuable.

6.2 References for Introduction

Vasquez, K. (2023). *What to do when language impedes science?* C&EN. https://cen.acs.org/careers/diversity/language-impedes-science/101/i38

6.3 Copyright

Reprinted from “Deng, J. M.; Flynn, A. B. “I Am Working 24/7, but I Can’t Translate That to You”: The Barriers, Strategies, and Needed Supports Reported by Chemistry Trainees from English-as-an-Additional Language Backgrounds. *J. Chem. Educ.* **2023**, *100* (4), 1523–1536. doi.org/10.1021/acs.jchemed.2c01063” with permission from the American Chemical Society and Division of Chemical Education, Inc. Copyright © 2023 American Chemical Society and Division of Chemical Education, Inc.

JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

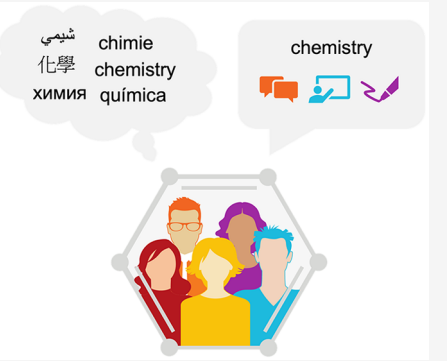
"I Am Working 24/7, but I Can't Translate That to You": The Barriers, Strategies, and Needed Supports Reported by Chemistry Trainees from English-as-an-Additional Language Backgrounds

Jacky M. Deng and Alison B. Flynn*

Cite This: J. Chem. Educ. 2023, 100, 1523−1536 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Having a shared language such as English facilitates the globalization of science, but it can also limit science by posing barriers for those who have learned English-as-an-additional language (Eng+). Research on the experiences of Eng+ graduate students and postdoctoral researchers ("trainees") is needed as our scientific community becomes more diverse and continues working toward improving equity and inclusion; without such research, we risk losing out on perspectives and expertise that might otherwise strengthen scientific progress. We investigated three research questions: (1) What are Eng+ chemistry trainees' language experiences before entering English-dominant research programs? (2) What language barriers have trainees faced in learning, communicating, and doing chemistry? and (3) What strategies and supports have been helpful or needed? In focus groups and interviews with 18 Eng+ trainees, participants described having diverse language experiences before entering English-dominant research programs and faced challenges working in an English research environment. Participants reported avoiding interactions with colleagues, exacerbated feelings of imposter syndrome, needing extra time to learn or communicate, and not being able to fully express their knowledge. Participants expressed an ongoing desire to learn English and described various independent strategies to navigate challenges. All participants believed that research supervisors and institutions needed to play a larger role in supporting Eng+ trainees, such as supervisors who offer personal and professional support, and institutional supports that provide chemistry-specific knowledge-building and networking opportunities. Based on the findings, we recommend approaches that supervisors and institutions can enact that may improve linguistic equity in the chemical sciences.



KEYWORDS: *equity and inclusion in chemistry, trainees, highly qualified personnel, graduate student education, international students in chemistry, english-as-an-additional language*

Downloaded via UNIV OF OTTAWA on April 18, 2024 at 21:32:41 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

LITERATURE REVIEW: CHEMISTRY TRAINEES COME FROM A DIVERSITY OF LANGUAGE BACKGROUNDS

Linguistic diversity in chemistry is increasing. There are approximately 1,900,000 international students studying in North America, the majority of whom come from nations where English is not a primary language (e.g., China, India, South Korea, Saudi Arabia).[1−4] In Canada, international students make up over a quarter of the total postsecondary student population.[2] At the graduate student level, 38% of science and engineering graduate students are temporary visa holders.[5] Many laboratories and scientific events rely on English, even in countries like Germany, Sweden, France, Switzerland, and Israel.[6] English has become the common language of science used across a diversity of language backgrounds, raising questions about the effects of language on those conducting scientific research.

Having a shared language like English facilitates the globalization of science but can also limit science, as many scientists learn English-as-an-additional language (Eng+) and may not have the English communication skills to fully participate in the research environment.[7] A study on Eng+ biologists in Colombia found that these scientists faced significant barriers to publishing and presenting their research in English, including additional time and financial costs.[8] Eng+ individuals can also struggle with linguistic insecurity (feelings of anxiety related to their use of English language) and accent discrimination.[9−14] Within the scientific publication system, structural inequities exist that favor English publications and journals.[6−8,15,16]

With ongoing efforts to promote and support diversity in the chemical sciences, the effects of English as an international

Received: October 31, 2022
Revised: March 2, 2023
Published: March 23, 2023



© 2023 American Chemical Society and Division of Chemical Education, Inc.

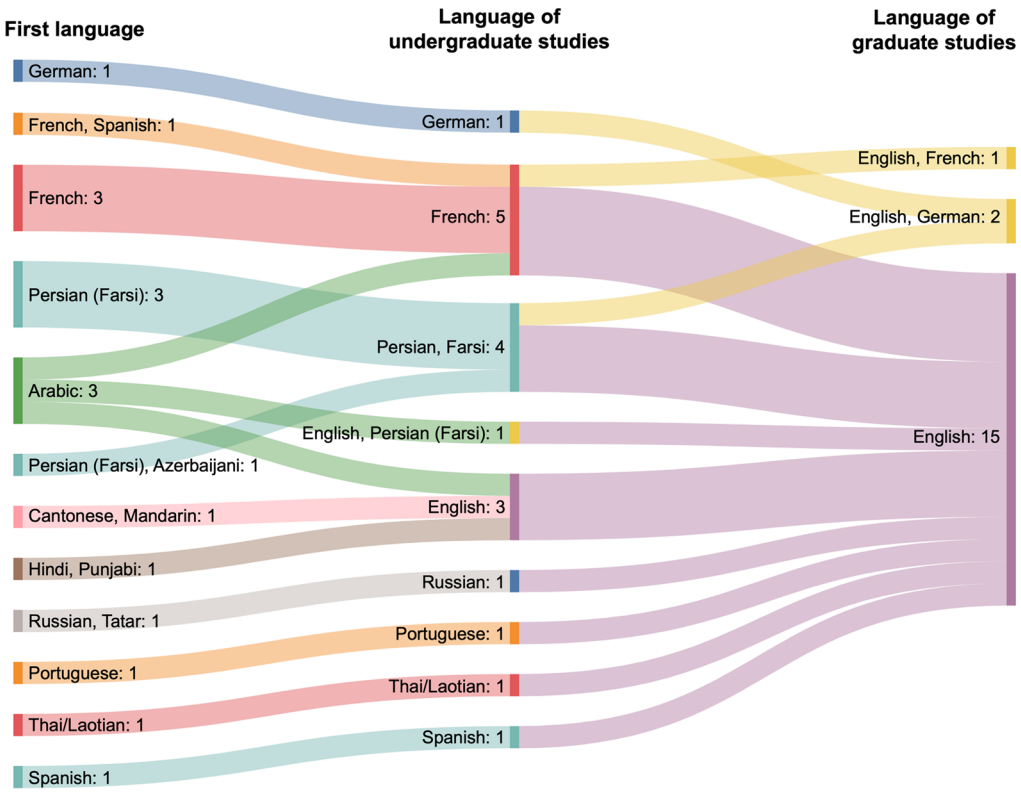


Figure 1. Participants' language backgrounds, specifically their first language, language of undergraduate studies, and language of graduate studies. N = 18.

language on a chemistry community with rich linguistic diversity must be considered, as language dictates how we **learn**, **communicate**, and **do** science.[14,17−19] Deeper understanding of the role of language in science may also lead to a more holistic understanding of the experiences of racialized chemists, creating pathways toward equity, inclusion, and justice in science more broadly.[20−22]

In this study, we investigated how language shapes Eng+ chemistry graduate students' and postdoctoral researchers' ("trainees") experiences in chemistry research, guided by the following research questions:

1. What are Eng+ trainees' language experiences prior to working and learning in English chemistry research programs?
2. What barriers have Eng+ trainees faced while learning, communicating, and doing chemistry?
3. What strategies and supports have Eng+ trainees found to be helpful and/or needed?

Though there has been previous research on Eng+ learners in chemistry education,[23,24] this study is the first to focus on Eng+ learners at the graduate and postdoctoral level. Many trainees come from international contexts and represent the next generation of scientists, playing a key role in sustaining and strengthening our future economies, environment, security, and health, but whose educational experiences are relatively understudied.[4,5] Through a better understanding of Eng+ trainees' experiences, our community can be better equipped to support Eng+ chemists through their studies and careers, making chemistry more equitable and inclusive, and enhancing scientific and economic progress.[7,16,25]

POSITIONALITY STATEMENT

Research is affected by the identities and perspectives of the researchers conducting the study.[26,27] This section describes the roles of our research team and the identities that may impact how we approach this work. Author J.M.D. (he/him) is a second-generation Chinese immigrant born in Canada. He considers English to be his first language but grew up in a Cantonese-speaking home. Author J.M.D. is also a chemistry graduate student at the same institution as the participants of this study and is actively involved in equity and inclusion initiatives at this institution. These two aspects of J.M.D.'s identity motivated him to design and carry out this study. Author A.B.F. (she/her) was born in Ottawa, Canada in a family with a Francophone father and English-speaking mother, with English as the dominant language in the home. She learned French in school. Both authors work in a bilingual environment (English, French).

Having J.M.D. lead the design, data collection, and analysis for the study was essential for this work; alongside critical perspectives from the literature and the participants, J.M.D. drew on his personal experiences and identities as a second-generation immigrant and chemistry graduate student when designing the study and its research questions, interviewing participants, analyzing the data, and developing this manuscript. Throughout the study, we examined participants' experiences with asset-oriented framing, focusing on how systems of power (e.g., the expectation to use English in science) shape the experiences of participants (Eng+ chemistry trainees), as well as how these systems might change to better serve members from diverse language backgrounds.[10,28−31] For example, we use the term "English-as-an-additional language" in lieu of other terms like "English-as-a-second language" and "English language

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

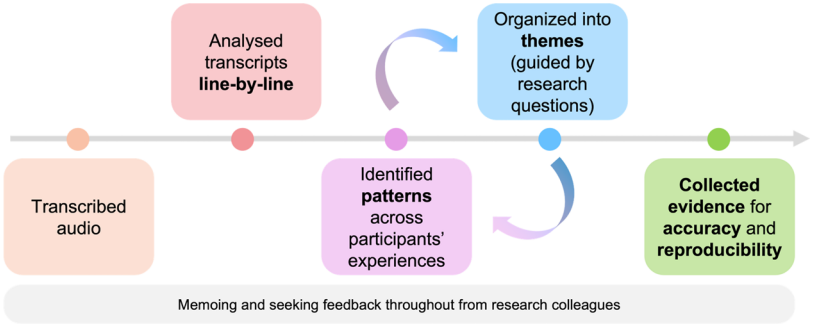


Figure 2. Overview of qualitative analysis.

learner" to better acknowledge the experiences, identities, and skills/knowledge Eng+ trainees *already* have before entering classrooms, research environments, etc.[32,33] We also deliberately pursued ways to honor participants' experiences and account for our own biases and assumptions (e.g., member-checking with participants).[34]

METHODS

We conducted this study at a bilingual (English, French) public research university in Canada. The bilingual context provided us with an opportunity to recruit participants from a diverse range of language backgrounds, as international students from 155 countries compose 17% of the student population at this institution.[35,36] We received ethics approval through the institution's Research Ethics Board (#H-01-22-7584).

The study sample (N = 18) was composed of 16 graduate students and 2 postdoctoral researchers (PDRs), collectively referred to as "trainees" throughout. Participants were involved in research across a range of chemistry subdisciplines (e.g., flow chemistry, biochemistry, organometallic, computational, organic), academic stages (3 MSc, 13 PhD, 2 PDR), and language backgrounds (Figure 1). Of the participants, 16 identified as international and 2 completed their undergraduate education in Québec, Canada, in which French is the dominant language. Participants were offered $25 gift cards as compensation.

We addressed the research questions via qualitative focus groups and interviews (each participant's choice), allowing participants to provide rich, detailed accounts of their experiences.[37] A total of three focus groups (each with three participants, average of 90 min) and nine interviews (average of 60 min) were conducted. All sessions were conducted and recorded virtually on Zoom due to the COVID-19 pandemic. Full focus group and interview protocols and questionnaires are provided in the SI.

We analyzed the focus group and interview data using qualitative methods (Figure 2, details in SI). First, we transcribed the audio recordings from the sessions verbatim to generate transcripts. This was followed by line-by-line coding of the transcript data. The goal of line-by-line coding is to break down the data into discrete parts (ideas, perspectives, etc.) to facilitate examining and comparing the data.[38,39] We focused on identifying distinct ideas within the transcripts, ranging in length from sentence fragments to statements up to four sentences long. After we identified initial codes from line-by-line coding, we pattern coded to identify broader categories. With these categories, the final stage of analysis involved organizing the data in terms of the research questions, allowing us to identify overarching findings/themes aligned with the research questions. Qualitative coding is an iterative process—we consulted relevant literature throughout the process and critically discussed our analysis and themes with each other and other members of our research group.[40] We also engaged in member-checking and inter-rater reliability procedures to maximize the validity and reliability of the research findings.[38,39] We invited participants to provide preferred pseudonyms to be used when communicating this research (including this article).

RESULTS AND DISCUSSION

This section describes and discusses the main findings uncovered from the analysis, organized by research question. Summaries of major findings (with exemplar quotes) for each research question are provided in Tables 1−4 and Figure 3.

RQ1: Eng+ Trainees Arrive at English Institutions with a Diversity of Language Experiences

Eng+ trainees arrive at English-speaking institutions with a variety of English histories and proficiencies. The 18 participants each varied in terms of how/where/when they learned English and the extent to which English was used as the language of instruction and learning for science/chemistry. Participants felt that their prior educations did not adequately prepare them for conducting research in English. Participants also expressed differing perspectives on the expectation to use English in science.

Having Limited or Sporadic Opportunities to Learn English. Most participants described having only *limited* opportunities to learn English during their prior education. Participants described having little-to-no exposure to English in their formal education and primarily learned English independently to complete standardized language tests to meet application requirements for graduate school (Table 1a). Other participants described *sporadic* opportunities to learn and use English, including attending supplemental courses on English or learning English independently for specific day-to-day tasks (e.g., through reading books, watching television, listening to music) (Table 1b). Participants who came from *limited* and *sporadic* experiences described how they learned English separate from scientific content (i.e., learning general grammar, structure, etc., but not *using* English when learning, communicating, or doing science), which has been shown to be less effective for language learning than immersive experiences.[41−43] Only one participant described an *immersive* English learning experience in which English was used daily in school (Table 1c).

Varying Perspectives on the Expectation to Use English in Science. Participants held diverse perspectives on the need to learn English to participate in science. Some stated that it is *fair to have to learn English*, viewing that the positives of having a common language in science outweighed individual

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

Table 1. Overview of Participants' Prior English Experiences and Perspectives on Learning English for Science

Extent of prior English learning experiences	Sample responses
(a) Limited	Marjan: "Iran is 100% Farsi. We don't have any English courses. We have English as just a course in our school. But most of the students that are learning English, it's not because of the school, it's because their parents are sending them outside for private class. And we are learning English just by ourselves and doing the TOEFL exam just for with ourselves. We don't have any school with the English language."
(b) Sporadic	Dave: "I did like a three-month course at British Council, and then the rest of learned basically on magazines, books, television series, things like that…it was a very short course. What I learned, I really learned it by force because I needed to work and to work I need to teach and I need to know English."
(c) Immersive	Abbad: "I went to an international school, so everything I learned was in English. We did have Arabic classes as language classes, but everything we took was in English."
Perspectives on English expectation	**Sample responses**
(d) Fair to have to learn and use English in science	Jennifer: "It's fair... I mean, it's a task to do, but we chose to do it. I mean, to be here you better have the ability or will to do it. Because we're not forced to come here. It was your plan, it was your idea to come here and work here. So I think it's more upon the person, because you cannot as an individual try to change the world and say 'hey everyone start speaking my language'. If they speak your language, it's the same thing for them. It's no different in any way."
	Xena: "I mean, it's sad, but I think it's fair, because we are living in an English-speaking country and English is the language of science."
	Marjan: "When we are coming here as new students, we are thinking that 'No, it's not fair.' It's really difficult and more pressure. But in the future…we need to be very well [in English] to compete for jobs. And because of that, we need to improve it. It's difficult, it's hard, but we have to do it."
(e) Issue is not learning English, but access to resources/supports to learn	Marat: "I mean, we would need to learn one language to communicate in any case and English does not seem like a bad choice, because it's not as hard as some other languages. I think it's normal. I would just like to have some course about how to properly write and maybe how to speak more freely. Something to help me improve my English speaking skills."
	Daniella: "English schools and exchanges are expensive, and a lot of families cannot afford them. Although one universal language is a way to increase communication among researchers, it is also a way to exclude people from less economically favored backgrounds."
	Alicia: "I think also that it's fair to learn English for science, because science is international, and we have to get the information from all over the world. So it's very important to have just one language, so we understand each other…I think one thing that I was searching [for] when I came here was a place where I can talk freely and practice English without having this feeling that people would judge me because I don't speak good English. Sometimes you are afraid of speaking freely because you are afraid of making mistakes."
(f) Loss of scientific knowledge and expertise	Sunny: "A lot of successful scientists or researchers in different countries, they didn't have a lot of opportunity as English-native speakers because they were lacking the opportunity, for example, for networking because they couldn't speak English. They had to do in their own country and that limited a lot of their opportunities. That is not fair."
	Daniella: "Um, no. I don't think it's fair. In a point that should be able to succeed in science, you have to learn English. And I was reading an article, I think [it] was Nature Science, about the difficulty of researchers from other countries that are not English speakers to have their work recognized specifically because they don't speak English. And this is really challenging because we see really good researchers that don't have English [as a] first language and they don't compete with other researchers that have English. I don't think this is fair, because we're missing a lot on the quality of researchers."

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

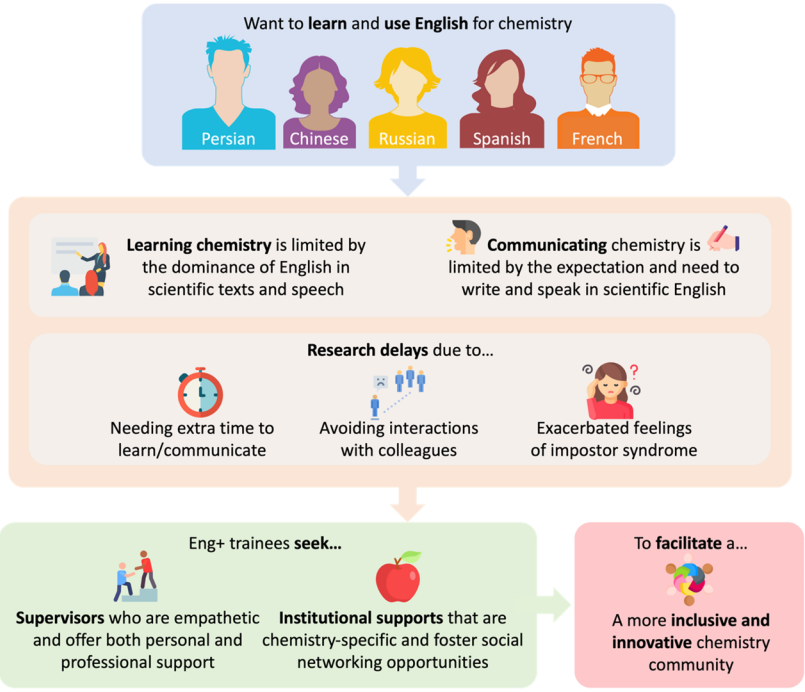


Figure 3. Eng+ trainees come from a variety of language backgrounds and described challenges associated with learning, communicating, and doing chemistry in English, as well as how supervisors and institutions might better support them in navigating challenges.

challenges (Jennifer, Table 1d). Others were less enthused about having to learn English, viewing the expectation as fair and offering expanded career and social prospects, while acknowledging that it posed individual challenges (Xena, Marjan, Table 1d)

For other participants, they stated that the *issue is not learning English, but access to resources/supports* (Table 1e). These participants focused on the limited access to opportunities and resources to learn English in both their prior and present educational contexts. By limiting access to English supports to those with sufficient socioeconomic capital, these participants stated that learning English becomes dependent on life factors largely beyond the control of learners, resulting in consequences for individuals' personal and professional development in science.[28,44]

One group of participants focused on the *loss of scientific knowledge and expertise* (Table 1f). These participants focused on the social and professional costs imposed upon Eng+ chemists in contexts where research was conducted in non-English languages, with implications for whose research is heard, read, and valued within the scientific community. Specifically, participants questioned whether it was fair for scientific quality and expertise to be associated with communicating and doing science in English. Echoing previous literature, participants stated that by making it more challenging for Eng+ chemists to broadly communicate their knowledge, the scientific community loses out on the expertise and knowledge of scientists from diverse linguistic/cultural backgrounds.[7]

RQ2: Eng+ Chemistry Trainees Face Barriers in Learning, Communicating, and Doing Chemistry in English

All 18 participants described challenges with **learning** (listening and reading) and **communicating** (speaking and writing) in English, with consequences to their ability to **do** chemistry.

Learning Chemistry Is Limited by the Dominance of English in Scientific Texts and Speech (Table 2a–d). All participants expressed having challenges with understanding scientific content in English, such as when reading articles, listening to presentations, following research discussions during group meetings and conferences, and day-to-day discussions with labmates and other colleagues.

For both reading and listening, participants described challenges with understanding scientific content from texts and presentations due to (1) being unfamiliar with English structure, grammar, and vocabulary, and (2) being unfamiliar with scientific terms in English (Table 2a,b). Scientific writing and jargon is a challenge for trainees in general,[45,46] but Eng+ trainees face additional challenges due to the need to learn both general/conversational English and chemistry-specific vocabulary.[17,24]

When listening to others in presentations, day-to-day research discussions, etc., participants described challenges associated with understanding different accents (Table 2c). Challenges with accents were exacerbated when participants struggled to keep up with the pace and/or volume of content in discussions (Table 2d). When others speak quickly and/or with an accent, participants noted that it could be very challenging for them to follow the discussion, causing them to miss out on content or feel unable to respond in a conversation.

Participants emphasized that their troubles with understanding content in scientific texts and presentations was due to the content being *presented* in English, and not the content itself. Kira stated that when she read scientific texts in English, she often missed key details that hindered her research and learning (Kira, Table 2a). Daniella described how reading and understanding scientific texts in English was more challenging than her first language (Portuguese), feeling like the latter "just flows" compared to English (Daniella, Table 2a). Daniella's use of both Portuguese and English is aligned with the phenomenon of code-switching, which describes how Eng+ learners can alternate between two or more languages within a given situation to accomplish a specific goal (i.e., reading).[23,47,48]

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

Table 2. Overview of Participants' Barriers with Learning and Communicating Chemistry in English

Learning chemistry is limited by the dominance of English in scientific texts and speech	Sample responses
(a) Struggling to learn chemistry because unfamiliar with English structure, grammar, and/or vocabulary	Marjan: "When, for example, when I was focusing more on the reading, I was just missing the topic. I was focusing on the words—what's the meaning of, for example, this one and this one. I was reading the whole paragraph, but once the paragraph finished, I didn't understand what's happening with the scientific [content]. And I had to re-read again to understand, 'Okay, what's the concept behind these things?'"
	Daniella: "Sometimes when I'm reading something—specifically if I'm learning something for a course—I read it and okay, I understand in English, [but] sometimes I have a problem understanding. I have to read it again, specifically, when there's something more complicated…And whenever I take the same subject, and I read it in Portuguese, it just flows. So I noticed that even though like I can read in English, I noticed whenever I need to learn something, sometimes I just put [it] in Google Translate. I read it, it flows. I don't know what it is, it just gets into my brain easier. So I noticed this. Or sometimes I have to read things really fast. I have a paper to read really fast. I do read faster in my first language. So I just put in Google Translate and quickly go through the words."
	Kira; "Before when I used to read an article, I had this joy exploring science. And every time I read something, it was like 'Oh wow! This is nice.' But now it's 'Oh, how did I miss this? How could I not see this before?'"
(b) Struggling to learn chemistry because unfamiliar with English scientific jargon	Abbad: "I remember opening my first paper, and as I was reading, [I] was like [expletive]! All these words don't make any sense. Like I speak English and I do understand English, but these words are not English, English. These are scientific words, just scientific jargon."
(c) Struggling to understand different accents	Paula: "[My labmates] describe it as a bit more like a Southern USA accent. And sometimes he's [labmate] talking and he says something and I'm like, I didn't get it because I don't know that accent… So it makes interaction with him a bit more difficult just because sometimes it's in my head like, 'Oh, we're going to talk? What if I don't understand? How do I react?'".
(d) Struggling to follow speech due to the pace and/or volume of content	Paul: "Sometimes [my supervisor] comes to help me or to see what I'm doing in the lab. We're doing something, and then he goes through something fast with me and it's like 'yeah, you just got to do this and this and this.' Some stuff that they think is super easy…And in my head I'm just trying to process everything he's saying, but he goes through it so fast."
	Alicia: "My main challenge, it was almost always the listening skills. Because it's just the way I am. It's hard for me to… I'm slower when I listen to things. It takes more time for me to process information. And when it was in English, because I told you that it was hard for me to understand presentation, in terms of the research in the first two years, it was kind of hard, because I needed to work twice harder to get the presentations, or the courses I took to try to really understand what the prof was saying, especially when it was graduates courses. I struggled. I struggled a lot because I had Anglophone teachers, professors, and the accent was not easy to follow and the technical subtleties were also hard to follow. So that was one of the hardest part of my PhD, [the] first two years."
Communicating chemistry is limited by the expectation to write and speak in academic English	**Sample responses**
(e) Struggling to use English structure, grammar, and/or vocabulary	Dave: "[Writing has] been a bit hard to be honest. I could definitely write longer texts in Spanish. Both papers—I have two papers—and it was a bit of a hell to write them because all my ideas definitely are simpler when I write in English. And I find it a bit more difficult to like create connections between the sentences. I also tend to write longer sentences because in Spanish, those are allowed and in English, they tend to be more concise. So I tend to overdo it. And sometimes I try to shorten things too much. And then I need to explain more."
(f) Struggling to communicate in scientific English	Xena: "Sometimes I want to make the manuscript look more professional, so I'm trying to use like more sophisticated words that sometimes aren't used in everyday conversational life. So that would take a little bit more energy compared to writing in my own language."
(g) Concerned with being viewed as less knowledgeable by others	Marat: "But with some native speakers, basically, you're thinking that they see you as the kind of dumb people because you basically struggling with pronunciation and all that stuff. You want for them to have a great opinion about you. Even if you know chemistry, if you kind of struggling with language, it kind of diminishes your value."
(h) Struggling to produce immediate responses	Alicia: "If I don't have a preparation beforehand or if I'm talking as I do the presentation to the people to the person [improvising], I would not be clear. And especially when people ask me questions, answering with something logical—introduction, development and conclusion—it's very hard. So if I don't have a preparation beforehand, the presentation is just not good."

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

Table 3. Participants Believed That Learning and Communication Challenges Contributed to Research Delays

Learning and communication challenges contribute to delays in research	Sample responses
(a) Resource implications	Marat: "It doesn't occur with like experimental work that we are doing. It more hurts when you have to write texts, like way longer time to write, and it kind of slows your pace."
	Daniella: "I think writing is challenging even in my first language. At least for me, it's my nightmare. Doesn't matter which language I'm speaking. But of course, writing English adds another layer of challenge on writing, specific for scientific writing because it requires a specific vocabulary, requires more grammar. So for me, yes, I do think it's challenging and I have a feeling it takes for me longer than the other students to write. So to write a normal paper, I see students sitting down in one month, they write the paper and I'm trying, it's been six months. I can't [laughs] I just can't. No paragraphs are coming out. So for me, it's what I call a nightmare—writing."
	Abita: "And so I've had a hard time with like, writing papers—like it takes me longer than like an English speaking person, because I'm mostly translating in my head and like writing it out as if I was speaking it and saying it."
	Jennifer: "I think at the end of the day, the people who know the language better have a better grip on writing, and takes them less time. I think, like the people who do not have it as a first language takes more time. Or sometimes, even I would say that sometimes I use the paraphrasing tool to put my words correctly."
(b) Social and psychological implications	Alica: "All the courses I took were actually really important for my research. They were linked. So because I couldn't understand a lot of the things that were said during the courses, it drained my energy in a way that I felt a little bit... you know, the impostor syndrome? Where you're like 'oh everyone here understands everything that the prof is saying, but no, not me.' So I didn't feel okay. I was emotionally feeling bad. So that impacted my research because I felt like I was not in the right place. It was emotionally draining."
	Kira: "I feel like I can do way more if things were better in the terms of like language, being able to communicate in the lab and, you know, skip some of the steps that I have to do myself [by asking] somebody else [who] can do it with me. But I don't know how to approach them."

Communicating Chemistry Is Limited by the Expectation to Write and Speak in Academic English (Table 2e–h). Communicating chemistry in English was described by several participants to be their most pervasive challenge, including writing about their research, presenting their research at conferences/meetings, communicating in research and day-to-day discussions with colleagues and supervisors, and networking at conferences.

When writing in English, participants described challenges (1) using English with the appropriate structure, grammar, and vocabulary, and (2) expressing their thoughts in scientific English (Table 2e,f). Similar to code-switching when reading, several participants described writing as involving a constant process of translating back-and-forth between their first language and English. Writing in English was also described by some participants as a stressful and cognitively demanding task, stating that they "avoid writing as much as possible" because "they knew [they were] bad it".

Speaking in English was also a challenge for participants, including speaking in both academic contexts (research presentations, discussions, etc.) and social contexts (day-to-day interactions with supervisors, colleagues, etc.). Some participants did not view research presentations as challenging as more spontaneous research discussions because they could prepare their presentations in detail beforehand, though they felt they often spent more time and effort preparing for presentations compared to Eng1st colleagues. Participants viewed the question periods of research presentations to be highly stressful due to the fear of misunderstanding audience questions and/or having to produce an immediate response to questions (Table 2h). Some stated that these fears stemmed from prior experiences in which they misunderstood an audience question.

When communicating, participants described how their thoughts and ideas often felt like they went through a "filter", with their ability to communicate ideas clearly or with the desired complexity tied to their ability to do so in English. Some described feelings of linguistic insecurity and fear of accent discrimination from colleagues (Table 2g).[9−11,49,50] Both these factors contributed to beliefs that their Eng1st colleagues and peers viewed them as less knowledgeable.[8,13]

Learning and Communication Contribute to Delays in Research (Table 3a,b). Though participants did not view language as a barrier in the practical aspects of doing chemistry (e.g., preparing and conducting experiments), participants noted that challenges learning and communicating chemistry in English resulted in research delays. Writing for Eng+ trainees is a longer process compared to Eng1st colleagues because (1) it takes longer to express thoughts and ideas clearly in English because of the additional time needed to use strategies such as

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

Table 4. Participants' Recommendations for Supervisors and Institutions to Support Eng+ Trainees

Trainees demonstrate resilience and a growth mindset in face of challenges	Sample responses
(a) Independently seeking out resources and support	Marjan: "Don't be ashamed. And just try to join in on events as much as possible. It doesn't matter that what is the topic—any conference, any seminars, any events, or just anything that students are trying to gather—just join. And find friends; being alone, it's not a good way."
	Daniella: "So I had a couple of students say 'Oh it was so great listening to you like, where are you from? I'm from a different country, too. I felt comfortable. I felt so comfortable with you talking that I felt comfortable making question for you.' For me this was really nice, and a lot of people that identified with me when I was presenting."
(b) Adopting a growth mindset	Alicia: "Working on self is more important than what others tell you. If you try to improve into yourself, because it's a language, you will not get it in a day. And it comes with practice. So if you do the practice on your part, I guess with time it improves. It's not something that difficult that you cannot learn. So I think there should be a will to learn and if the person has that, it can be done easily. It doesn't matter. Yes, if you're in the positive environment, it's encouraging. But if you're not, then you do your own part and improve."
	Nima: "We can't change our supervisor, our society, or people, but we can easily change ourselves. We can be positive about that. Don't give up, and try more and more to be able to actually improve our English."
Supervisors who provide both psychosocial and professional support	**Sample responses**
(c) Patience and empathy	Kira: "Compared to my colleagues, I'm way behind. And I receive comments from my supervisor, like, 'you have to work harder'. I am working hard. I am working 24/7, but I can't translate that to you. It's hard to translate to you that I am working 24/7. Because for someone whose English is first language, working 24/7 will achieve more than I will…I got the right education. [In Iran], I worked hard, I had time, and I practiced, so I achieved some things. But now in terms of daily communications, I need time, I need practice, I need someone to believe that I am capable. In order to do that, I need the power from inside. I need to believe in myself. I lost the self-esteem that I had."
	Paula: "[My supervisor] speaks a bit of French, so he will always try to talk a bit of French to me…You know, just a quick exchange in French, just to make me feel that 'I know you're French and I'm seeing that you're French and I want you to feel accepted.' He's not forgetting that if science doesn't work, that doesn't mean that the student doesn't work. He never forgot that. And that's great. It's really good that someone doesn't forget that that your success is not based solely on your work."
	Marjan: "In my mind, the supervisors who are taking the international students should be familiar that these students have a totally different background in science, in language, and in culture. And it takes time to adapt it to a new situation. And yeah, and they should be a little more patient and just give them more time. Because it's really challenging and more stressful."
	Daniella: "Whenever [my supervisor is] giving feedback, you feel that he's giving it with passion, with improvement in mind…Whenever he correcting us, you feel that he's not there to throw you down. He's just adding, he always starts with 'Oh, this is so good. The way you write is so good. It's just at the end that we should just add something like this…' So he always has this passion to throw you up instead of down."
	Jennifer: "So in the initial days, I was a bit hesitant to meet my supervisor on a frequent basis, because I was like, 'Oh my God, what will I go and say?' At the time, he figured out that I was nervous to reach out to him, and so he called me once to his office and said, 'Hey, I see you are nervous. And there's nothing to be nervous about. Talk to me.'"

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

Table 4. continued

Supervisors who provide both psychosocial and professional support	Sample responses
(d) Providing trainees with resources and supports	Xena: "So when I started, it was hard, so I started asking my lab mates for their feedback and they were nice enough to proofread what I was writing. And then also, my supervisor, took his time to go through the mistakes I had and explain it to me. And he also introduced a grammar book for scientific writing. So that was helpful."
	Jennifer: "And in order to take help me in a way, so my supervisor gave me a book, which was called, 'Introduction to Writing on Chemistry'. And like the basics of writing on chemistry, he gave it to me like last week or 10 days ago when I showed him my first small draft of my paper. And he gave me the book in the vision that he himself reads that book most of the time. So he said, because it's my first time writing, he would like me to read it in my free time, and I have been following that."
	J.M.D.: "What about for writing?" Dave: "Writing? I don't think there is much of a difference just because my supervisor, she tends to make like a standardized way for when we write papers... It's more organized, I guess." J.M.D.: "Do you like having that structure?" Dave: "I think so, I think so. It definitely helps to be more efficient. And to focus more on the ideas."
Institutional supports that are discipline-specific and foster community	**Sample responses**
(e) Discipline-specific	Cecilia: "They send us to the English writing center. I found that when the TA or the center don't have a strong science background, then they couldn't give like a lot helpful advice when we get to things that are more academic or more science. So it would be helpful... for example, I'm doing biochemistry, so if there was anyone like have like chemistry background or biology backgrounds that can do the proofreading, then I'm sure she can give more advice on the writing other than the general one."
(f) Communities of Eng+ trainees	Daniella: "As soon as I finished [my presentation], we went to the posters. And I had a couple of students say 'Oh it was so great listening to you like, where are you from? I'm from a different country, too. I felt comfortable. I felt so comfortable with you talking that I felt comfortable making question for you.' For me this was really nice, and a lot of people that identified with me when I was presenting."
	Dave: "The other students have the same issues: pronunciation, when writing. So it's almost like a safe space to talk with people who also struggle with English because it's not their mother tongue."
	Alicia: "I think one thing that I was searching for when I came here was a place where I can talk freely and practice English without having this feeling that people would judge me because I don't speak good English. I don't know if it's the case, but sometimes you are afraid of speaking freely because you are afraid of making mistakes."

code-switching,[23,48] (2) it requires more revisions and feedback from colleagues, and (3) they may avoid writing because they associate English with negative feelings and experiences (Table 3a). Though participants often engaged in independent strategies to compensate for language challenges (e.g., Grammarly, Google Translate, proactively seeking feedback from peers), they also noted that these strategies required extra time and steps that they believed Eng1st trainees did not have to rely on as frequently.

Communication challenges resulted in social barriers that made it harder for participants to establish effective working relationships with their colleagues and supervisors. They described situations in which they actively avoided asking for help from their colleagues to be viewed as competent and to compensate for how others might view their language challenges. Xena stated: "I think that [Eng+ trainees] need to excel in science. So if I say, one side is science, the other part is communication, the average should be similar. So if I'm less OK in communication, then I should be better in science." Others like Kira were unsure how to ask for support from their colleagues because they were unaware of how to do so in ways that were aligned with linguistic norms and power dynamics, leading to negative implications for their research productivity (Kira, Table 3b).

Learning and communication challenges left participants feeling like outsiders in science (i.e., impostor syndrome), making research more emotionally and mentally draining to do (Alicia, Table 3b). Impostor syndrome is a pervasive challenge

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

for trainees in general and can stem from not feeling a sense of community and affiliation—essential for developing a scientific identity. Feelings of impostor syndrome may be exacerbated for Eng+ trainees as being a member of the scientific community requires proficiency in English.[51−53]

Trainees Demonstrated Resilience and a Growth Mindset in Face of Challenges (Table 4a,b). Participants adopted various independent skills and strategies in response to the barriers they faced learning, communicating, and doing chemistry in English. These skills and strategies included code-switching to learn and communicate ideas in English, attending extra courses, workshops, and seminars to familiarize themselves with English, using translation and other language support technologies (e.g., Grammarly, Google Translate, DeepL) to support their writing/presentations, and proactively asking for feedback on their communications from colleagues and supervisors.

Participants also independently sought out communities of Eng+ trainees that not only allowed them learn English alongside other learners, but also build support networks with other Eng+ trainees. These communities allowed trainees to navigate their common challenges by sharing strategies, motivating each other, and providing each other with psychosocial support.[54,55] When asked what advice she would have for other Eng+ trainees, Marjan focused on the importance of seeking out social events to find friends and establish a support system (Marjan, Table 4a). Daniella recalled a moment when other Eng+ trainees saw her present at a conference, expressing to her that they felt inspired and reassured hearing another Eng+ trainee speak (Daniella, Table 4a).

Participants also responded to challenges by adopting a growth mindset: a belief their abilities and qualities can be developed and cultivated through dedication and hard work.[56,57] Alicia viewed her challenges with English as part of the learning process, believing that her English would improve with practice and time (Alicia, Table 4b). Believing that he could not change how his supervisor and environment supported him, Nima stated that he and other Eng+ trainees like him can still "change ourselves" to improve their English (Nima, Table 4b).

Though qualities like grit, resilience, and perseverance are admirable, the associated statements came from participants who felt they had minimal support from their supervisors and institutions. In other words, these qualities are responses to environments and systems that leave Eng+ trainees to fend for themselves.[58] Should we expect Eng+ trainees to be grittier and more resilient? Or should we instead direct our efforts to better support Eng+ trainees as colleagues, supervisors, and institutions, especially as our community strives to create more equitable, inclusive, and diverse research and learning environments?[55,59,60] The following section details themes related to strategies and supports that Eng+ trainees viewed as helpful and/or needed in navigating language challenges in chemistry.

RQ3: What Strategies and Supports Have Eng+ Chemistry Trainees Found Helpful and/or Needed?

Supervisors Who Provide Both Psychosocial and Professional Support (Table 4c,d). Participants who described their supervisors as impatient and unempathetic viewed their language barriers as independent challenges they had to navigate alone. In some cases, participants stated their supervisors not only left them feeling alone, but also exacerbated their language challenges by expecting them to be more productive to compensate for their language challenges. These instances left participants frustrated with their supervisors' lack of empathy toward their language difficulties, not only negatively affecting their research, but also their confidence as a scientist (Kira, Table 4c).

In contrast, participants who described having positive mentor relationships with their supervisor appreciated how their supervisor offered both psychosocial support (e.g., psychological and emotional support, role modeling) and professional support (e.g., career guidance, professional development opportunities).[61] Supervisors who demonstrated *patience and empathy* included those who proactively invited trainees to reach out for psychological and emotional support (Table 4c). Other examples included supervisors who reassured trainees as they navigated challenges, such as when trainees were worried about adapting to their new culture/environment, comprehensive exams, etc., and supervisors who made efforts to learn more about trainees' linguistic and cultural backgrounds (Table 4c). Other work has suggested that it may be helpful to provide Eng+ learners with extra time to analyze and respond to information (e.g., questions, feedback) due to Eng+ learners reliance on strategies like code-switching.[23]

Participants also appreciated supervisors who *provided trainees with resources and supports* to develop their English and other chemistry-specific communication skills (Table 4d). This included providing trainees with resources to support their language skills, like sharing writing/presentation guides and templates, and providing clear and constructive feedback on manuscripts, presentations, etc. For example, Xena and Jennifer's supervisors both lent them books on scientific writing to help them strengthen their communication skills. Dave appreciated how their supervisor provided a writing template to provide all members of their group with a scaffold to organize their writing.

Institutional Supports That Are Discipline-Specific and Foster Community (Table 4e,f). Institutional supports are broadly defined in this work as resources offered by departments, faculties, institutions, etc. Participants desired institutional supports that were specific to chemistry (i.e., discipline-specific) (Table 4e). For example, participants found their institution's academic writing center unhelpful because of the center's inability to offer constructive feedback on chemistry-specific writing. Participants suggested establishing a chemistry- or science-specific writing center in which trainees could meet with other trainees (either Eng+ or Eng1st) to receive support with writing and presenting in science/chemistry. Others felt that chemistry-specific communication courses could also help with becoming accustomed to the norms of communicating chemistry (e.g., writing abstracts for chemistry conferences/manuscripts, preparing chemistry presentations, communicating research to a lay audience).[62−65] Preferences for discipline-specific resources align with recommendations to embed language learning *within* meaningful content and contexts.[43,66]

Participants also expressed a desire for community spaces that provided them with opportunities to not only practice their English with others, but also develop relationships with others in the community, including Eng+ peers (Table 4f). Examples of these included conversation groups or trainee/student associations and networks. Social belonging interventions have been shown to positively impact Eng+ learners' outcomes in postsecondary education.[67] Such interventions provide learners with formal spaces to connect with others like them, allowing

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

members to draw on a community when navigating shared challenges.[54,55]

IMPLICATIONS

This work is the first to investigate the experiences of Eng+ trainees in science and its findings present broad implications, including for mentor−mentee relationships in science.

Eng+ trainees are a diverse group who enter science with a range of English language experiences and proficiencies, and most participants in this study described having only limited or sporadic opportunities to learn English in their previous educational contexts. Though Eng+ trainees may be required to pass English language assessments prior to learning and working in English-dominant institutions, pathways to accomplishing this are heterogeneous and may not adequately prepare trainees for completing day-to-day scientific work in English.[68] These opportunity gaps can result in Eng+ trainees facing a multitude of barriers when trying to learn and communicate chemistry in English, with consequences that impact their ability to productively do chemistry. We must identify equitable ways to support Eng+ trainees in science, especially as science becomes increasingly globalized and thus increasingly dependent on English. Therefore, we share Eng+ trainees' suggested actions for supervisors and institutions, providing readers with a starting point to create more equitable and inclusive research and learning environments for those from diverse language backgrounds.

Suggested Actions for Research Supervisors

Participants preferred supervisors who demonstrated *patience and empathy* toward their experiences, and *provided trainees with resources and supports* to develop their English and chemistry-specific communication skills. For diverse trainees to thrive in science, supervisors must create safe and trusting environments that are respectful of trainees' backgrounds and experiences.[31,69−72] We have summarized specific actions supervisors might enact based on our findings, as well as other recommendations from the literature on scientific mentorship.[4,61,73] We note that the findings and recommendations provided are based on a qualitative study; we cannot comment on the relative prevalence of the findings and recommendations, nor do we have evidence related to the efficacy of recommendations—more research is needed.

- **Build trust**. Trainees face significant mental, financial, and social stressors, and these are often amplified for those from diverse linguistic and cultural backgrounds.[74−77] Effective mentors recognize this and work to establish trust with trainees through regular meetings and reassuring trainees that they can (to the best of their ability) provide support/resources. These characteristics are especially important in cultivating cross-cultural mentor−mentee relationships that foster and support diverse trainees.[78−80]
- **Proactively offer resources and supports to Eng+ trainees to support their development**, including writing/presentation resources and templates, providing consistent and constructive feedback on writing/presentations, encouraging trainees to pursue extracurricular opportunities (e.g., English courses), directing trainees to mental health supports, connecting them with diverse mentors, etc. Such resources might be included on a group Web site, on-boarding guide for new trainees, etc.

Suggested Actions for Institutions

In addition to mentors' beliefs and practices, educational change requires consideration of contextual factors, such as the departmental and institutional context.[81,82] Participants recommended implementing institutional supports that were (1) chemistry-specific and (2) allowed them to foster connections with other Eng+ trainees/scientists. Based on these findings, recommended actions for institutions include:

- **Establishing chemistry-specific communication resources** to provide trainees with communication supports that are discipline-specific. These could take the form of discipline-specific communication courses (e.g., Communication in Chemistry) that provide trainees with opportunities to use English for learning and communicating in chemistry.[43] Other resources could include writing centers embedded within departments or faculties. These could be led by fellow trainees (Eng+ or Eng1st) to promote the development of mentorship skills among trainees and foster social connections.
- **Creating community spaces for Eng+ trainees** that allow them to share strategies, skills, and initiatives with each other. At the departmental or institutional level, community spaces might be realized through Eng+ trainee associations/networks that connect Eng+ trainees with peers/mentors from similar backgrounds.[54,55,83] There may also be opportunities to create national/international scientific societies that cater specifically to the needs and interests of Eng+ scientists.[84]

CONCLUSIONS

English may be the language of science, but it is not the first language of many scientists and trainees. We investigated how the expectation to use English in chemistry shapes the research experiences of 18 Eng+ chemistry trainees. Through focus groups and interviews, we found that the participants arrived at an English-dominant research institution with primarily *limited* or *sporadic* English experiences, but with a desire to learn English to participate in science. Participants described a variety of challenges associated with *learning* (understanding papers, listening to presentations, etc.) and *communicating* (writing papers, participating in research discussions, etc.) chemistry in English, which resulted in negative consequences for participants' abilities to *do* chemistry (spending extra time reading/writing, fear of asking for help with research, etc.). Aligned with their desire to learn English, trainees independently sought out and adopted a variety of strategies to navigate challenges, such as attending extracurricular language courses, using translation technologies, and creating learning communities; these strategies are indicative of a community of learners that demonstrates resilience, capable of viewing challenges as opportunities. Trainees felt that supervisors play an important role in both the scientific and linguistic development of Eng+ trainees, feeling more comfortable and supported when working with supervisors who (1) demonstrated patience and empathy toward Eng+ trainees' experiences in science and (2) proactively supported trainee development through strategies like providing constructive feedback and sharing chemistry-specific language resources with trainees (e.g., research writing/presentation guides and templates, chemistry-specific communication courses/workshops). Participants' institutional recommendations included creating chemistry-specific communication

https://doi.org/10.1021/acs.jchemed.2c01063
J. Chem. Educ. 2023, 100, 1523−1536

resources (e.g., courses, writing centers) and spaces for Eng+ community building (e.g., conversation spaces).

This study provides us with a better understanding of Eng+ trainees' experiences, including suggesting some specific actions to better support Eng+ chemists during their studies and careers. Such work is imperative in fostering a scientific community that is more equitable, inclusive, and innovative.[7,16,25]

LIMITATIONS

Because we chose to focus exclusively on the role of language in this work, we did not identify how different aspects of participants' identities may have intersected with and contributed to their experiences (e.g., race/ethnicity, gender, sex) nor did we distinguish between the different languages represented in the study sample. We also do not know how our bilingual institutional context may have shaped participants' experiences (e.g., if French-first participants might have different experiences compared to participants with no French background at our English/French institution). Our previous work with undergraduate science students suggests that one's preference for using English and/or French when communicating chemical ideas is associated with their language background (e.g., extent to which they used a language in school vs home).[14] However, more research is needed to determine whether this is also true for Eng+ trainees, especially because of the broader expectation in the scientific community to communicate and work in English. Future research could investigate how bilingual institutions shape the experiences of Eng+ trainees, if at all.

We did not explicitly ask participants if they experienced communication disabilities, although we asked participants to report whether they had a disability in the demographic questionnaire (3 yes, 15 no) and no discussions about communication disabilities arose during discussions. Furthermore, although we engaged in a process of member-checking to verify whether the findings and themes aligned with participants' experiences, we did not observe participants directly and our findings are limited to what participants were able to describe verbally to us in English. Future studies are needed to explore the limitations described above and related issues.

Lastly, because all our participants identified as Eng+, we cannot comment on or compare against the perspectives and experiences of Eng1st participants, including whether Eng1st participants would also find the language learning resources and recommendations suggested by the Eng+ participants in this study to be useful (or not) in developing their chemistry-specific communication skills. Previous research on language learning (outside of chemistry education) suggests that Eng+ learners may benefit more from tailored instruction focused on: writing genre expectations and social experiences as writers; grammar and conventional patterns;[85−88] enacting feedback;[89,90] applying strategies effective for Eng1st writers;[91] and text complexity, cohesion, clause length, and organization.[92] There have also been several projects reported in this journal on chemistry-specific communication resources—for example, writing courses and workshops in which learners receive writing instruction and/or work toward a developing a review article, grant proposal, etc.[62−65] However, none of these articles explicitly distinguish or compare between Eng1st and Eng+ participants and only one focused explicitly on Eng+ graduate-level participants. Future research distinguishing between the perspectives and experiences of Eng1st and Eng+ trainees may be useful in developing strategies and resources to support all and/or specific groups of learners (depending on project goals).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.2c01063.

Qualitative codebook with full descriptions of themes and additional examples; Supplemental text on the theoretical framework guiding this research; Supplemental text on the methods used in this research; Supplemental data from participants' responses to the pre-session and post-session questionnaires (PDF) (DOCX)
Pre-session questionnaire (PDF)
Post-session questionnaire (PDF)
Interview focus group protocol (PDF)

AUTHOR INFORMATION

Corresponding Author

Alison B. Flynn − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada;* orcid.org/0000-0002-9240-1287; Email: alison.flynn@uottawa.ca

Author

Jacky M. Deng − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada;* orcid.org/0000-0002-2633-6624

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.2c01063

Funding

This research was supported in part by the Social Sciences and Humanities Research Council (SSHRC) through a Vanier Canadian Graduate Scholarship and through an Insight Grant.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Sarah McPhedran for serving as the second rater when evaluating inter-rater reliability in this study. We thank other members of the Flynn Research Group for their helpful, critical feedback at various stages of this study. Lastly, we thank the participants for sharing their stories with us; we hope this work can contribute to efforts to support them and others from similar backgrounds. Author J.M.D. dedicates this work to his parents.

ABBREVIATIONS

Eng+, English-as-an-additional language; Eng1st, English-as-a-first language; PDR, Postdoctoral researcher; SI, Supporting Information.

REFERENCES

(1) *Project Atlas: A Quick Look at Global Mobility Trends*; Institute of International Education, 2020.

(2) *The State of Post-Secondary Education in Canada*; Higher Education Strategy Associates, 2020.

(3) Crossman, E.; Lu, Y.; Hou, F. International Students as a Source of Labour Supply: Engagement in the Labour Market after Graduation. *Econ. Soc. Rep.* **2022**, *2* (3), 1−5.

(4) National Academies of Science Engineering and Medicine. *Graduate STEM Education for the 21st Century*; Leshner, A., Scherer, L., Eds.; National Academies Press: Washington, D.C., 2018. DOI: 10.17226/25038.

(5) *2019 Doctorate Recipients from U.S. Universities*; National Science Foundation, 2020; Vol. *NSF 21-308*.
(6) *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*; Ammon, U., Ed.; De Gruyter Mouton, 2011. DOI: 10.1515/9783110869484.
(7) Amano, T.; González-Varo, J. P.; Sutherland, W. J. Languages Are Still a Major Barrier to Global Science. *PLoS Biol.* **2016**, *14* (12), No. e2000933.
(8) Ramírez-Castañeda, V. Disadvantages in Preparing and Publishing Scientific Papers Caused by the Dominance of the English Language in Science: The Case of Colombian Researchers in Biological Sciences. *PLoS One* **2020**, *15* (9), e0238372.
(9) Subtirelu, N. C. She Does Have an Accent but: Race and Language Ideology in Students' Evaluations of Mathematics Instructors on RateMyProfessors.Com. *Lang. Soc.* **2015**, *44* (1), 35−62.
(10) Lippi-Green, R. Accent, Standard Language Ideology, and Discriminatory Pretext in the Courts. *Lang. Soc.* **1994**, *23* (2), 163−198.
(11) Chin, W. Y. Linguistic Profiling in Education: How Accent Bias Denis Equal Educational Opportunites to Students of Color. *Scholar* **2010**, 356−384.
(12) Rubin, D. L. Nonlanguage Factors Affecting Undergraduates' Judgments of Non-native English-speaking Teaching Assistants. *Res. High. Educ.* **1992**, *33*, 511−531.
(13) Cheng, Y.; Dharwadkar, S.; Sheridan, V.; Rojas, C. R.; Bosch Grau, M.; Gordin, M. Science's Language Barrier. *Nature* **2019**, *570*, 265.
(14) Deng, J. M.; Rahmani, M.; Flynn, A. B. The Role of Language in Students' Justifications of Scientific Phenomena. *Int. J. Sci. Educ.* **2022**, *44* (13), 2131−2151.
(15) Amano, T.; Rios Rojas, C.; Boum, Y.; Calvo, M.; Misra, B. B. Ten Tips for Overcoming Language Barriers in Science. *Nat. Hum. Behav.* **2021**, *5*, 1119.
(16) Angulo, E.; Diagne, C.; Ballesteros-Mejia, L.; Adamjy, T.; Ahmed, D. A.; Akulov, E.; Banerjee, A. K.; Capinha, C.; Dia, C. A. K. M.; Dobigny, G.; Duboscq-Carra, V. G.; Golivets, M.; Haubrock, P. J.; Heringer, G.; Kirichenko, N.; Kourantidou, M.; Liu, C.; Nuñez, M. A.; Renault, D.; Roiz, D.; Taheri, A.; Verbrugge, L. N. H.; Watari, Y.; Xiong, W.; Courchamp, F. Non-English Languages Enrich Scientific Knowledge: The Example of Economic Costs of Biological Invasions. *Sci. Total Environ.* **2021**, *775*, 144441.
(17) Lee, O.; Fradd, S. H. Science for All, Including Students from Non-English-Language Backgrounds. *Educ. Res.* **1998**, *27* (4), 12−21.
(18) Lee, E. N.; Orgill, M.; Kardash, C. Supporting English Language Learners in College Science Classroom: Insights from Chemistry Students. *Multicult. Educ.* **2020**, *27*, 25−32.
(19) Elsherief, H.; Masson, M. *Every Teacher Is a Language Teacher*; University of Ottawa Second Language Education Cohort (cL2c), 2020; Vol. *1*. DOI: 10.20381/an3d-6k16.
(20) Museus, S. D.; Iftikar, J. An Asian Critical Theory (AsianCrit) Framework. *Asian Am. Students High. Educ.* **2013**, 18−29.
(21) Iftikar, J. S.; Museus, S. D. On the Utility of Asian Critical (AsianCrit) Theory in the Field of Education. *Int. J. Qual. Stud. Educ.* **2018**, *31* (10), 935−949.
(22) Valdes, F. Latina/o Ethnicities, Critical Race Theory, and Post-Identity Politics in Postmodern Legal Culture: From Practices to Possibilites Foreword. *Berkeley La Raza Law J.* **1996**, *9* (1), 1115593.
(23) Flores, A.; Smith, K. C. Spanish-Speaking English Language Learners' Experiences in High School Chemistry Education. *J. Chem. Educ.* **2013**, *90* (2), 152−158.
(24) Lee, E. N.; Orgill, M. Toward Equitable Assessment of English Language Learners in General Chemistry: Identifying Supportive Features in Assessment Items. *J. Chem. Educ.* **2022**, *99*, 35.
(25) Hofstra, B.; Kulkarni, V. V.; Munoz-Najar Galvez, S.; He, B.; Jurafsky, D.; McFarland, D. A. The Diversity−Innovation Paradox in Science. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (17), 9284−9291.
(26) Howe, M. E.; Kim, M. M.; Pazicni, S. Graduate Student Women's Perceptions of Faculty Careers: The Critical Role of Departmental Values and Support in Career Choice. *J. Am. Chem. Soc. Au* **2022**, *2*, 1443.
(27) Curran, M.; Randall, A. K. Positionality Statements. https://onlinelibrary.wiley.com/pb-assets/assets/14756811/Positionality-Statements-1621354517813.pdf (accessed on May 22, 2022). DOI: 10.1177/1745691620927709.
(28) Ladson-Billings, G. From the Achievement Gap to the Education Debt: Understanding Achievement in U. S. Schools. *Educ. Res.* **2006**, *35* (7), 3−12.
(29) Shukla, S. Y.; Theobald, E. J.; Abraham, J. K.; Price, R. M. Reframing Educational Outcomes: Moving beyond Achievement Gaps. *CBE Life Sci. Educ.* **2022**, *21* (2), 1−11.
(30) León, M. Standard Language Ideologies, World Englishes, and English Language Teaching: An Overview. *Arizona Work. Pap.* **2018**, *25*, 44−62.
(31) Emdin, C. *For White Folks Who Teach in the Hood... and the Rest of Y'all Too: Reality Pedagogy and Urban Education*; Beacon Press: Boston, MA, 2016.
(32) González-Howard, M.; Suárez, E. Retiring the Term English Language Learners: Moving toward Linguistic Justice through Asset-Oriented Framing. *J. Res. Sci. Teach.* **2021**, *58*, 749−752.
(33) Grapin, S. E. Toward Asset-Oriented and Definitionally Clear Terminology: A Comment on González-Howard and Suárez (2021). *J. Res. Sci. Teach.* **2021**, *58* (5), 753−755.
(34) Creswell, J. W.; Miller, D. L. Determining Validity in Qualitative Inquiry. *Theory Pract.* **2000**, *39* (3), 124−130.
(35) University of Ottawa. Quick Facts 2020. https://www.uottawa.ca/institutional-research-planning/resources/facts-figures/quick-facts (accessed on May 10, 2022).
(36) *Programs and Admissions*; University of Ottawa, 2021.
(37) Braun, V.; Clarke, V. *Successful Qualitative Research: A Practical Guide for Beginners*; Sage Publications Inc.: London, 2013.
(38) Saldaña, J. *The Coding Manual for Qualitative Researchers*, 3rd ed.; Sage Publications Inc.: Thousand Oaks, CA, 2016.
(39) Bazeley, P.; Jackson, K. *Qualitative Data Analysis with NVivo*; Sage Publications Inc.: Thousand Oaks, CA, 2013; Vol. *40*. DOI: 10.1080/02607476.2013.866724.
(40) Kember, D.; Ha, T. S.; Lam, B. H.; Lee, A.; Ng, S.; Yan, L.; Yum, J. C. K. The Diverse Role of the Critical Friend in Supporting Educational Action Research Projects. *Educ. Action Res.* **1997**, *5* (3), 463−481.
(41) The National Academy of Sciences; National Academies of Science Engineering and Medicine; The National Academy of Sciences. *English Learners in STEM Subjects: Transforming Classrooms, Schools, and Lives*; Francis, D.; Stephens, A., Eds.; The National Academies Press: Washington, D.C., 2018. DOI: 10.17226/25182.
(42) Lee, O. Asset-Oriented Framing of Science and Language Learning with Multilingual Learners. *J. Res. Sci. Teach.* **2021**, *58* (7), 1073−1079.
(43) McKay, S. L. Teaching Materials for English as an International Language. In *Principles and Practices of Teaching English as an International Language*; Matsuda, A., Ed.; Multilingual Matters: Bristol, 2012; pp 70−83.
(44) Savva, M. International Schools as Gateways to the Intercultural Development of North-American Teachers. *J. Res. Int. Educ.* **2013**, *12* (3), 214−227.
(45) Pain, E. How to (Seriously) Read a Scientific Paper. *Science (80-.).* **2016**, 1−7.
(46) Carey, M. A.; Steiner, K. L.; Petri, W. A. Ten Simple Rules for Reading a Scientific Paper. *PLoS Comput. Biol.* **2020**, *16* (7), e1008032.
(47) Hansson, A. The Meaning of Mathematics Instruction in Multilingual Classrooms: Analyzing the Importance of Responsibility for Learning. *Educ. Stud. Math.* **2012**, *81* (1), 103−125.
(48) Lin, A. M. Y. Code-Switching in the Classroom: Research Paradigms and Approaches. *Res. Methods Lang. Educ.* **2016**, *10*, 1−15.
(49) Pawlas, E.; Paradowski, M. B. Misunderstandings in Communicating in English as a Lingua Franca: Causes, Prevention, and Remediation Strategies. In *The Role of Languages in Intercultural Communication*; Koutny, I.; Stria, I.; Farris, M., Eds.; Poznań, 2020; pp 99−120.

(50) Nguyen, H. T. What Role Do Race, Ethnicity, Language and Gender Play in the Teaching Profession? *Race Ethn. Educ.* **2012**, *15* (5), 653−681.
(51) Brown, B. A. Discursive Identity: Assimilation into the Culture of Science and Its Implications for Minority Students. *J. Res. Sci. Teach.* **2004**, *41* (8), 810−834.
(52) Brown, B. A.; Ryoo, K. Teaching Science as a Language: A "Content-First" Approach to Science Teaching. *J. Res. Sci. Teach.* **2008**, *45* (5), 529−553.
(53) Vincent-Ruz, P.; Schunn, C. D. The Nature of Science Identity and Its Role as the Driver of Student Choices. *Int. J. STEM Educ.* **2018**, DOI: 10.1186/s40594-018-0140-5.
(54) Yosso, T. J. Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth. *Race Ethn. Educ.* **2005**, *8* (1), 69−91.
(55) Deng, J. M.; McMunn, L. E.; Oakley, M. S.; Dang, H. T.; Rodriguez, R. S. Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities. *J. Chem. Educ.* **2022**, *99* (1), 373−382.
(56) Flynn, A. B.; Campbell-Brown, E.; O'Connor, E. K.; Walsh, E.; O'Hagan, F.; Richard, G.; Roy, K. What Is a Mindset? In *Growth & Goals: A Course-Integrated Module to Better Equip Students with Learning Skills*; eCampus Ontario, 2022.
(57) Dweck, C. S. *Mindset: The New Psychology of Success*; Gildan Media Corp., 2007.
(58) Mcgee, E. O.; Stovall, D. Reimagining Critical Race Theory in Education: Mental Health, Healing, and the Pathway to Liberatory Praxis. *Educ. Theory* **2015**, *65* (5), 491−511.
(59) Wilson-Kennedy, Z. S.; Payton-Stewart, F.; Winfield, L. L. Toward Intentional Diversity, Equity, and Respect in Chemistry Research and Practice. *J. Chem. Educ.* **2020**, *97* (8), 2041−2044.
(60) Sanford, M. Equity and Inclusion in the Chemical Sciences Requires Actions Not Just Words. *J. Am. Chem. Soc.* **2020**, *142*, 11317−11318.
(61) National Academies of Science Engineering and Medicine. *The Science of Effective Mentorship in STEMM*; Byars-Winston, A.; Dahlberg, M. L., Eds.; National Academies Press: Washington, D.C., 2019.
(62) Stewart, A. F.; Williams, A. L.; Lofgreen, J. E.; Edgar, L. J. G.; Hoch, L. B.; Dicks, A. P. Chemistry Writing Instruction and Training: Implementing a Comprehensive Approach to Improving Student Communication Skills. *J. Chem. Educ.* **2016**, *93* (1), 86−92.
(63) Shufflebarger, A. M.; Scott, K. A. Chemistry Departments Should Facilitate Graduate-Level Second-Language Writing Instruction. *J. Chem. Educ.* **2020**, *97* (12), 4220−4224.
(64) Ogunsolu, O. O.; Wang, J. C.; Hanson, K. Writing a Review Article: A Graduate Level Writing Class. *J. Chem. Educ.* **2018**, *95* (5), 810−816.
(65) McCarthy, B. D.; Dempsey, J. L. Cultivating Advanced Technical Writing Skills through a Graduate-Level Course on Writing Research Proposals. *J. Chem. Educ.* **2017**, *94* (6), 696−702.
(66) Lee, O.; Stephens, A. English Learners in STEM Subjects: Contemporary Views on STEM Subjects and Language With English Learners. *Educ. Res.* **2020**, *49* (6), 426−432.
(67) LaCosse, J.; Canning, E. A.; Bowman, N. A.; Murphy, M. C.; Logel, C. A Social-Belonging Intervention Improves STEM Outcomes for Students Who Speak English as a Second Language. *Sci. Adv.* **2020**, *6* (40), 1−11.
(68) Hamid, M. O.; Hardy, I.; Reyes, V. Test-Takers' Perspectives on a Global Test of English: Questions of Fairness, Justice and Validity. *Lang. Test. Asia* **2019**, DOI: 10.1186/s40468-019-0092-9.
(69) White, K. N.; Vincent-Layton, K.; Villarreal, B. Equitable and Inclusive Practices Designed to Reduce Equity Gaps in Undergraduate Chemistry Courses. *J. Chem. Educ.* **2021**, *98* (2), 330−339.
(70) Braun, D. C.; Gormally, C.; Clark, M. D. The Deaf Mentoring Survey: A Community Cultural Wealth Framework for Measuring Mentoring Effectiveness with Underrepresented Students. *CBE Life Sci. Educ.* **2017**, *16* (1), 1−14.
(71) Denofrio, L. A.; Russell, B.; Lopatto, D.; Lu, Y. Mentoring: Linking Student Interests to Science Curricula. *Science (80-.).* **2007**, *318* (5858), 1872−1873.
(72) Mireles-Rios, R.; Garcia, N. M. What Would Your Ideal Graduate Mentoring Program Look like?: Latina/o Student Success in Higher Education. *J. Latinos Educ.* **2019**, *18* (4), 376−386.
(73) Tuma, T. T.; Adams, J. D.; Hultquist, B. C.; Dolan, E. L. The Dark Side of Development: A Systems Characterization of the Negative Mentoring Experiences of Doctoral Students. *CBE Life Sci. Educ.* **2021**, *20* (2), ar16.
(74) Pasic, M.; Eleftheriades, R.; Fiala, C. The Challenges and Mental Health Issues of Academic Trainees. *F1000Research* **2020**, *9*, 104.
(75) Hyun, J.; Quinn, B.; Madon, T.; Lustig, S. Mental Health Need, Awareness, and Use of Counseling Services among International Graduate Students. *J. Am. Coll. Heal.* **2007**, *56* (2), 109−118.
(76) Evans, T. M.; Bira, L.; Gastelum, J. B.; Weiss, L. T.; Vanderford, N. L. Evidence for a Mental Health Crisis in Graduate Education. *Nat. Biotechnol.* **2018**, *36* (3), 282−284.
(77) Levecque, K.; Anseel, F.; De Beuckelaer, A.; Van der Heyden, J.; Gisle, L. Work Organization and Mental Health Problems in PhD Students. *Res. Policy* **2017**, *46* (4), 868−879.
(78) Papadopoulos, I.; Lees, S. Developing Culturally Competent Researchers. *J. Adv. Nurs.* **2002**, *37* (3), 258−264.
(79) Campinha-Bacote, J. A Culturally Conscious Model of Mentoring. *Nurse Educ.* **2010**, *35* (3), 130−135.
(80) White-Lewis, D. K.; Romero, A. L.; Gutzwa, J. A.; Hurtado, S. Moving the Science Forward: Faculty Perceptions of Culturally Diverse Mentor Training Benefits, Challenges, and Support. *CBE Life Sci. Educ.* **2022**, *21* (1), 1−10.
(81) Szozda, A. R.; Bruyere, K.; Lee, H.; Mahaffy, P. G.; Flynn, A. B. Investigating Educators' Perspectives toward Systems Thinking in Chemistry Education from International Contexts. *J. Chem. Educ.* **2022**, *99* (7), 2474−2483.
(82) Gess-Newsome, J.; Southerland, S. A.; Johnston, A.; Woodbury, S. Educational Reform, Personal Practical Theories, and Dissatisfaction: The Anatomy of Change in College Science Teaching. *Am. Educ. Res. J.* **2003**, *40* (3), 731−767.
(83) Rogoff, B. Developing Understanding of the Idea of Communities of Learners. *Mind, Cult. Act.* **1994**, *1* (4), 209−229.
(84) Inside eLife. The Languages of Science—multilingual Science Communication Awareness. https://elifesciences.org/inside-elife/d4f67431/webinar-report-the-languages-of-science-multilingual-science-communication-awareness (accessed on September 7, 2022).
(85) Zhao, J. Native Speaker Advantage in Academic Writing? Conjunctive Realizations in EAP Writing by Four Groups of Writers. *Ampersand* **2017**, *4*, 47−57.
(86) Appel, R.; Szeib, A. Linking Adverbials in L2 English Academic Writing: L1-Related Differences. *System* **2018**, *78*, 115−129.
(87) Hinkel, E. Indirectness in L1 and L2 Academic Writing. *J. Pragmat.* **1997**, *27* (3), 361−386.
(88) Pan, F.; Reppen, R.; Biber, D. Comparing Patterns of L1 versus L2 English Academic Professionals: Lexical Bundles in Telecommunications Research Journals. *J. English Acad. Purp.* **2016**, *21*, 60−71.
(89) Farrokhi, F.; Sattarpour, S. The Effects of Direct Written Corrective Feedback on Improvement of Grammatical Accuracy of High- Proficient L2 Learners. *World J. Educ.* **2012**, *2* (2), 49−57.
(90) Cheng, A. Analyzing and Enacting Academic Criticism: The Case of an L2 Graduate Learner of Academic Writing. *J. Second Lang. Writ.* **2006**, *15* (4), 279−306.
(91) Guo, X.; Huang, L. S. Are L1 and L2 Strategies Transferable? An Exploration of the L1 and L2 Writing Strategies of Chinese Graduate Students. *Lang. Learn. J.* **2020**, *48* (6), 715−737.
(92) Eckstein, G.; Ferris, D. Comparing L1 and L2 Texts and Writers in First-Year Composition. *TESOL Q.* **2018**, *52* (1), 137−162.

Chapter 7: The experiences of chemistry researchers from diverse language backgrounds

7.1 Introduction to the research presented in this chapter

Our research in Chapter 6 was a qualitative study; though the goal provided us with a depth of understanding into the Eng+ chemistry trainee experience, it did not provide a breadth of understanding. For example, we generated an understanding of what specific challenges Eng+ trainees might face, and the reasons for them, but we did not have data to suggest which challenges were more or less prevalent and/or generalizable to the broader population. Such data would be valuable in understanding the breadth of specific challenges or resource preferences and help us determine what supports and resources could be prioritized when considering ways to support Eng+ scientists. Guided by our qualitative findings from Chapter 6, Chapter 7 describes a study focused on the following research questions: For chemistry researchers from diverse language backgrounds: (1) How does the need to use English in science impact their research experiences and perceived productivity? (2) What resources and strategies do they use to navigate language barriers during their research? (3) In what ways are they using generative artificial intelligence, if at all? The chapter details how we developed and administered a questionnaire, how we collected and analyzed data from the questionnaire, as well as our preliminary findings from quantitative and qualitative data analyses. It is the only non-published manuscript in my thesis, with our data analysis to be completed. Thank you to Makayla Dagenais for her contributions to the analysis of the data shown in Figure 4.

7.2 Introduction

7.2.1 English is the language of science, but not all scientists

English is the language we primarily use in science, but is not the first language of many scientists. Today, science is increasingly conducted around the world by scientists who learn English-as-an-additional language (Eng+). Having a shared language like English facilitates globalization of science but can also limit science, as many scientists learn English-as-an-additional language (Eng+) and may not have the English communication skills to fully participate in the research environment. A growing body of research has demonstrated that the dominance of English in science poses unique challenges for both Eng+ scientists, including publishing manuscripts, presenting at conferences, and applying to research funding (Amano et al., 2016, 2023; Deng & Flynn, 2023; Ramírez-Castañeda, 2020). Eng+ individuals can also struggle with linguistic insecurity (feelings of anxiety related to their use of language) and accent discrimination (Cheng et al., 2019; Chin, 2010; Deng et al., 2022; Foote, 2010; Lippi-Green, 1994; Subtirelu, 2015).

Much of the existing research at the intersection of language and science has been focused at the undergraduate and K-12 context. Eng+ students' achievement in STEM fields has been found to be associated with language ability, regardless of students' abilities to grasp and apply scientific concepts (Curtis & Millar, 1988; Maerten-Rivera et al., 2010; Solano-Flores & Trumbull, 2003). Eng+ students' written responses can score lower on evaluations than those of Eng1st students due to evaluators' implicit bias (Huang, 2008; Lindsey & Crusan, 2011; Milnes & Cheng, 2008) and students' challenges communicating their knowledge (Lyon et al., 2012; Swanson et al., 2014; Wolf et al., 2008). Several accommodations and resources to support Eng+ students have been reported, though many of these have been focused in undergraduate education. For example, providing extra time, linguistic modifications that simplify linguistic complexity in assessment items, providing bilingual glossaries/dictionaries. Institutional supports

such as social-belonging interventions have also been shown to be helpful for strengthening feelings of inclusion for English newcomers (LaCosse et al., 2020). However, many classroom-based existing interventions lack evidence to support their proposed impacts and may actually limit Eng+ students' success if applied with careful consideration of students' linguistic abilities and needs (Abedi, 2002; Abedi et al., 2020; Buxton et al., 2014).

There has been a growing body of research on the Eng+ experience in scientific research. Our previous research qualitatively investigated the experiences of Eng+ chemistry trainees ($N = 18$) (Deng & Flynn, 2023). Through focus groups and interviews, we learned that participants faced challenges in research due to the expectation and need to use English for communicating and learning in chemistry. These challenges reportedly had consequences on participants' research participation, due to: (1) the extra time required for tasks like reading and writing; (2) avoiding social interactions with peers, including asking for help in their research; and (3) exacerbated feelings of imposter syndrome. However, in facing these challenges, participants also discussed strategies and supports they found useful and/or needed. These discussions centred on the need for supervisors and colleagues who could provide both psychosocial and professional support, as well as discipline-specific institutional supports (*e.g.*, chemistry-specific writing centres).

Other work has been conducted in the field of biology on the continents of Europe, Asia, and South America. Less work has been conducted in North America, despite both Canada and the US serving as global leaders in scientific research and English-dominant regions of the world with many citizens and scientists from diverse language backgrounds (Dietrich & Hernandez, 2022; Government of Canada, 2022; National Science Board, 2021). Additionally, much of the existing research has focused on faculty members and/or researchers in industry/government (Amano et al., 2016, 2023; Ramírez-Castañeda, 2020). Research that includes the perspectives and experiences of trainees (graduate students and postdoctoral researchers) is essential, as an increasing number of trainees come from international contexts and represent the next generation of scientists, playing a key role in sustaining and strengthening our future economies,

environment, security, and health (National Academies of Science Engineering and Medicine, 2018; National Science Foundation, 2020). There have been few studies examining the effects of specific strategies to support researchers from diverse language backgrounds, with some demonstrating positive effects from interventions focused on sense of belonging (LaCosse et al., 2020) and graduate-level writing courses (Shufflebarger & Scott, 2020).

Linguistic equity is especially important in an age of linguistic diversity in science. There are 1,900,000 international students studying in North America, the majority of whom come from nations where English is not a primary language (*e.g.*, China, India, South Korea, Saudi Arabia) (Crossman et al., 2022; Higher Education Strategy Associates, 2020; Institute of International Education, 2020; National Academies of Science Engineering and Medicine, 2018). In Canada, international students make up over a quarter of the total post-secondary student population (Higher Education Strategy Associates, 2020). At the graduate student level, 38% of science and engineering graduate students are temporary visa holders (National Science Foundation, 2020). Many laboratories and scientific events rely on English, even in countries like Germany, Sweden, France, Switzerland, and Israel (Ammon, 2011).

Though a common language like English allows our global scientific community to exchange ideas and communicates, the challenges it poses for certain individuals can have negative downstream consequences for science and science communication. Neglecting the issues facing Eng+ scientists can result in increased attrition of students from a diversity of linguistic, cultural, and ethnic and backgrounds, and curbed development of key scientific skills (Abedi, 2002; Afitska & Heaton, 2019a; Allensworth, 2005; Buxton et al., 2014, 2015; Chin, 2010; Maerten-Rivera et al., 2010; McNeil, 2005; Pyburn et al., 2013; Solano-Flores & Trumbull, 2003; Sotelo-Dynega et al., 2013; The National Academy of Sciences et al., 2018; Wolf et al., 2008; Yore et al., 2004). We as a scientific community lose out on expertise and research if any researchers faces barriers to participating in science (Amano et al., 2016, 2021; Ammon, 2011; Angulo et al., 2021; Ramírez-Castañeda, 2020). With ongoing efforts to promote and support diversity in the

chemical sciences, the effects of English as an international language on a chemistry community with rich linguistic diversity must be considered, as language dictates how we **learn**, **communicate**, and **do** science (Deng et al., 2022; Elsherief & Masson, 2020; E. N. Lee et al., 2020; O. Lee & Fradd, 1998).

7.2.2 Generative artificial intelligence (GenAI) to support linguistic equity

The field of artificial intelligence (AI) has recently experienced unprecedented growth, ushering in a new era of transformative tools and possibilities across various industries. In particular, generative artificial intelligence (GenAI) tools built on powerful deep learning algorithms possess the capacity to receive and generate human-like text, enabling them to engage in contextually relevant conversations with users. The frontrunner of large language models (LLMs), ChatGPT, represents a significant breakthrough in natural language processing, heralding a new paradigm for educators and researchers.

Tools like ChatGPT have already brought forth complex discussions and debates within education and research. Concerns around academic integrity have led some institutions to ban students and researchers from using GenAI (Elsen-Rooney, 2023; Nature, 2023b). However, others have argued for the opposite; these AI tools exist, continue to develop, and are readily accessible to those with internet access; banning these tools risks foster ignorance while increasing the appeal and safety risks for users (Botten, 2023; Kirk, 2023; Roose, 2023). Therefore, it is vital for our scientific community to learn about these tools, including how people use them, so that we can ensure that all learners and researchers are able to use them effectively and ethically (Holmes, 2023; Wiley, 2023).

Thus far, much of the existing literature on GenAI in chemistry education and research has focused on ways to adapt and leverage tools like ChatGPT in the chemistry classroom (assessment, curriculum development, adaptive learning, *etc*.), including its limitations (Alasadi & Baiz, 2023; Clark, 2023; dos Santos, 2023; Emenike & Emenike, 2023; Fergus et al., 2023;

Gardner & Giordano, 2023; Humphry & Fuller, 2023; Talanquer, 2023). A recent article discussed the potential for GenAI tools to serve as a language support for those learning English (or any language) (Berdejo-Espinola & Amano, 2023), and guided by this article, our study explores how researchers from diverse language backgrounds use these tools during chemistry research.

7.3 Goals and research questions

This manuscript describes a study focused on the experiences of chemistry researchers from diverse language backgrounds. Our previous research (Chapter 6) was a qualitative study; we sought to explore a phenomenon, focusing on the *depth* of participants' experiences, attitudes, *etc*. However, the qualitative study had a smaller sample size, making it difficult to determine the *breadth* of participants' experiences, attitudes, *etc*. In other words, we did not learn how prevalent certain issues are for researchers across different situations or contexts, as well as what supports/resources could be prioritized to support researchers. Through a better understanding of Eng+ chemistry researchers' experiences, our community can be better equipped to support Eng+ chemists through their studies and careers, making chemistry more equitable and inclusive and enhancing scientific and economic progress (Amano et al., 2016, 2023; Angulo et al., 2021; Hofstra et al., 2020).

Guided by existing research literature (Amano et al., 2023; Deng & Flynn, 2023; Ramírez-Castañeda, 2020), as well as developments in generative artificial intelligence (Berdejo-Espinola & Amano, 2023), we developed and administered a questionnaire to collect and analyze quantitative and qualitative data focused on the following research questions:

For chemistry researchers from diverse language backgrounds:

1. How does the need to use English in science impact their research experiences and perceived productivity?

2. What resources and strategies do they use to navigate language barriers during their research?
3. In what ways are they using generative artificial intelligence, if at all?

7.4 Theoretical Framework: Learning, communicating, and doing science

The current project, as well as our previous work on this topic, have been guided by a theoretical framework focused on the role of language in learning, communicating, and doing science (O. Lee & Fradd, 1998). This framework argues that language plays an active role in science and science education; language is a means of doing science and constructing scientific understanding (Abedi, 2002; Afitska & Heaton, 2019b; Solano-Flores & Trumbull, 2003; The National Academy of Sciences et al., 2018; Wolf et al., 2008; Yore et al., 2004).

With this framing in mind, the dominance of English in science means Eng+ learners and scientists may face unique linguistic challenges with learning, communicating and doing science (O. Lee & Fradd, 1998). Knowing science means making meaning of scientific knowledge by connecting new information to previous knowledge. Language is foundational to learning, as it mediates how learners engage with scientific concepts and construct meaning. Learning new information can be a challenge for Eng+ students, as they may have learned prerequisite concepts in a language or culture different from English, making it more challenging to identify and establish coherent connections between new information and their previous knowledge. Communicating in science means communicating in the language of science: scientific English (Lemke, 1991). Learning and communicating in scientific English poses challenges for many learners, including Eng+ students who may have less experience with the traditional norms and patterns of scientific English (Deng & Flynn, 2023). For example, in contexts where scientists are asked to articulate an explanation or argument for a particular phenomenon, Eng+ individuals may engage in more cognitive operations compared to Eng+ students in order to both interpret English questions and then generate a response in English (Abedi, 2002, 2015; Abedi et al., 2004;

Abedi & Lord, 2001; Buxton et al., 2014; González-Howard & McNeill, 2016; E. N. Lee, 2018; E. N. Lee et al., 2020; E. N. Lee & Orgill, 2021; Solano-Flores & Trumbull, 2003). Doing science means engaging scientific practices, such as engaging in inquiry, proposing arguments and explanations, and interpreting and identifying data to make sense of phenomena (O. Lee et al., 2013; O. Lee & Fradd, 1998). Many of these practices require complex language functions and an understanding academic norms and traditions that require sufficient levels of English language proficiency within a given discipline.

7.5 Methods

7.5.1 Questionnaire development

To address the research questions with a breadth of data, we developed a questionnaire guided by established literature outlining best practices in instrument design (American Psychological Association et al., 2014; Barbera & VandenPlas, 2011; Stains, 2022).

We developed an initial draft of questionnaire items. These items were informed by a combination of sources, including (1) the research questions; (2) our previous research findings; (3) other research and theory on Eng+ experiences in science/academia (Amano et al., 2023; Berdejo-Espinola & Amano, 2023; LaCosse et al., 2020; E. N. Lee et al., 2020, 2023; O. Lee & Fradd, 1998; Ramírez-Castañeda, 2020); and (4) discussions and suggestions from colleagues. Questionnaire development is an iterative process—author JMD created an initial draft of items, shared these items with AF and group members for discussion and feedback, JMD revised the items, *etc*. The goal of this stage was to create a draft of questionnaire items that would address the overarching goals/questions guiding the project, before further refining the questionnaire through collecting validity evidence.

7.5.2 Overview of the questionnaire

The questionnaire is organized into two halves. The first half of the questionnaire focuses on participants' experiences using English for learning, communicating, and doing chemistry

research. Questions include Likert scale items in which participants are asked to rate their level of agreement or experience with statements about language and chemistry research. Participants were also asked to comment on their use of and perspectives on different language resources, including GenAI tools such as ChatGPT. In the second half of the questionnaire, participants were asked about their demographic background, including their career stage, race/ethnicity, disability, and country of work. These demographic data were collected to be used as co-variates in regression analyses, allowing us to focus on the role of language background while controlling for other demographic variables, aligned with our research questions.

7.5.3 Validity evidence

Validity refers to the degree to which evidence/theory support interpretations of data for proposed uses with a specific population (American Psychological Association et al., 2014; Cronbach, 1971; Kane, 2013; Messick, 1995). Collecting validity evidence is imperative in test development—without evidence for validity of the data obtained from a particular instrument in specific research context, there is uncertainty about what the instrument is actually measuring (American Psychological Association et al., 2014).

In some cases, such as this study, instruments may not yet exist for a particular goal, requiring instruments to be created and supported with appropriate validity evidence to determine whether they measure what they intend to measure. The effectiveness and suitability of an instrument in chemistry education research can vary depending on the goals and context of a particular evaluation. An instrument designed for a particular purpose in one context may need to be re-evaluated, re-configured, or re-designed to provide meaningful measurements in pursuit of new research goals and/or within a new context.

Validity can be comprised of evidence from instrument content, response process, internal structure, relationships with other variables, and consequences of instrument use (American Psychological Association et al., 2014; Messick, 1995). Each type of evidence can be collected

using specific and established research methods; readers interested in detailed definitions of each of these types of validity evidence and their collection methods are referred to the *Standards for Educational and Psychological Testing* (American Psychological Association et al., 2014).

For this project, we collected content validity evidence and response process validity evidence. Content validity evidence was important to collect, as we wanted to ensure we captured the breadth of experiences related to using English in scientific research, thereby establishing the questionnaire's relevance and comprehensiveness in assessing participants' experiences and perspectives. Response process validity evidence was also crucial; because respondents were likely to vary in terms of their English language proficiency, we wanted to ensure that the wording of questionnaire items was as clear and understandable as possible for participants.

Not every type of validity evidence is required for every research study because the relevance and importance of each type depend on the specific aims, context, and nature of the study. Prioritizing certain types of validity evidence over others allows researchers to focus their resources and efforts effectively, tailoring the validation process to the particular needs and constraints of their research (American Psychological Association et al., 2014; Barbera & VandenPlas, 2011). Other forms of validity evidence, such as internal structure and relationships with other variables, were not collected because they fell outside the scope of the study and required resources that outweighed the benefits of collecting these types of evidence.

7.5.4 Content validity evidence

Content validity refers to the extent to which a measurement instrument adequately covers the content domain it is intended to measure (American Psychological Association et al., 2014). In other words, content validity assesses whether the items or questions included in the instrument represent the full range of the topic, idea, construct, *etc.* being measured. In this study, content validity involved ensuring that the questionnaire adequately captured the various aspects of the

experiences of chemistry researchers from diverse language backgrounds in ways that aligned with our research questions.

To collect evidence for content validity, we contacted 15 researchers in the chemical sciences, inviting them to review the questionnaire (*i.e.*, expert review). We asked each researcher to: (1) complete the questionnaire once using an online link, taking note of any questions that they found confusing or had questions about; and (2) provide feedback/comments on a PDF version of the questionnaire, such as suggestions to make questions clearer, questions they think could be added/removed, *etc*. We collected content validity evidence in two rounds so that we could seek feedback in the first round, revise the questionnaire based on this feedback, and then use a second round of feedback to confirm whether our revisions addressed the initial suggestions, as well as identify any new concerns. Participants included trainees and faculty members, as well as a mix of Eng+ and Eng1st individuals. Drawing from this feedback, we modified the questionnaire to better reflect the language experiences of people from diverse language backgrounds, grounded in our research questions. For example, the first version of the questionnaire included an item about respondents' perspectives on their supervisors. Several faculty members who participated in the content validation process told us that this question was confusing, as they did not feel the question was relevant to their present experience. Responses from several trainees also suggested that this item was confusing. Considering both these perspectives, along with our research questions, we excluded the item in the final version of the questionnaire. The content validation process also revealed to us which questions to prioritize keeping. For example, several participants told us that the item "Feeling anxious or overwhelmed when expressing my thoughts in English" resonated strongly with their experience, motivating us to include this item in the final version of the questionnaire.

7.5.5 Response process validity evidence

We collected response process validity evidence for the questionnaire. Response process validity refers to the extent to which items are being interpreted by participants in the ways that the researchers intended (American Psychological Association et al., 2014; Deng et al., 2021).

To collect response process validity evidence, we conducted cognitive interviews, which are interviews to probe how respondents think about and respond to materials presented to them, with a specific emphasis on what barriers and pitfalls arise during their response process. We conducted cognitive interviews Zoom with three researchers, all of whom were Eng+ individuals. In each interview, the participant completed the questionnaire once using an online link. They then discussed with the interviewer any questions that they found confusing or unsure about when they went through the questionnaire (*i.e.*, retrospective approach) (Willis, 2011). Lastly, the interviewer would go through the questionnaire question-by-question with the participant (*i.e.*, concurrent approach), asking them questions such as "Can you repeat this question in your own words?", "How did you decide on your response?", and "What were you thinking about when you answered this question?" (Willis, 2011). The interviews were completed in two rounds, so that we could identify issues in the first round, revise, and receive feedback on revisions in the second round. Based on these interviews, we made minor adjustments to simplify the wording of some items. For example, the item "Concern that my manuscripts, proposals, conference abstracts, *etc*., are more likely to be rejected due to English ability" was viewed as too "wordy" by participants, and this item was revised to "Concern that my manuscripts are more likely to be rejected because of my English ability."

7.5.6 Research Ethics Board approval

We applied for and received ethics approval for this study through the University of Ottawa Research Ethics Board (REB) (Ethics File #H-01-22-7584).

7.5.7 Pilot study

We piloted the questionnaire at the University of Ottawa. The goal of the pilot study was to test collecting and analyzing data on a smaller scale before conducting the full-scale study. Pilot studies are commonly conducted to identify any issues with data collection or analysis before conducting full-scale studies (Malmqvist et al., 2019). We encountered no issues with data collection and analysis during the pilot process, and thus moved forward with the full-scale study.

7.5.8 Data collection

We invited chemistry researchers from any language background, country, career stage, and chemistry subdiscipline to participate. We recruited participants through the following strategies (Creswell, 2012):

- Direct emails to individuals in our academic networks and publicly available email lists of chemistry researchers (*e.g.,* DiversifyChem) (Diversify Chemistry, 2024)
- Professional listservs (*e.g.,* Canadian Council of University Chemistry Chairs (CCUCC))
- Social media (Twitter/X)
- Snowball recruiting: inviting people to share the study with others who might be interested

We received 225 responses to the questionnaire. These data were analyzed using a combination of qualitative and quantitative data analysis methods. The final questionnaire and recruitment posters are available in Appendix 4.

7.5.9 Qualitative data analysis

We used thematic analysis to analyze qualitative data (written responses) from the questionnaire (Bazeley & Jackson, 2013; Saldaña, 2016). Thematic analysis involves systematically categorizing and analyzing (*i.e.*, coding) textual or visual data to identify themes, patterns, and

concepts, allowing researchers to derive meaning and insights from qualitative data by organizing it into manageable units for interpretation and analysis.

The qualitative data were imported into the online whiteboard tool, *Miro* (RealtimeBoard Inc., 2024). We began by organizing data into initial categories guided by a learning, communicating, doing framework (O. Lee & Fradd, 1998). For example, if a participant's response included discussion of their experiences writing or resources they used to support their writing, this response would categorized under "communicating". After this process (known as deductive coding), we began to inductively code the data based on common ideas and topics expressed between responses. Inductive coding allowed us to code data that did not fit clearly into the deductive options (*e.g.*, feeling "worried about false information generated by GenAI"), as well as to identify more granular subcategories within the deductive categories. Once we identified an initial set of codes and categories, we pattern coded to identify broader categories. With these categories, the final stage of analysis involved organizing the data in terms of the research questions, allowing us to identify overarching findings/themes aligned with the research questions. Qualitative coding is an iterative process—we consulted relevant literature throughout the process and critically discussed our analysis and themes with each other and other members of our research group (Kember et al., 1997).

7.5.10 Quantitative data analysis

Quantitative data were analyzed using a combination of Microsoft Excel and SPSS. Descriptive statistics were processed in Excel. We used SPSS to conduct statistical tests, such as *t*-tests, chi-squared tests of independence, and multiple regression analyses (Field, 2014).

A *t*-test is a statistical test used to determine if there is a significant difference between the means of two groups. It assesses whether the means of the two groups are statistically different from each other, taking into account the variability of the data and the sample sizes. *t*-tests were chosen to assess whether there are significant differences in the experiences of chemists from

Eng+ vs. Eng1st backgrounds, guided by our research questions. *t*-tests are type of parametric test, meaning that they require normally distributed data to produce accurate results and reliable statistical inference (Field, 2014). In our study, *t*-tests are appropriate for comparing means between samples of $n = 99$ (Eng1st) and $n = 126$ (Eng+) because, according to the Central Limit Theorem, as sample sizes increase, the sampling distribution of the sample mean approaches a normal distribution regardless of the shape of the population distribution (Kwak & Kim, 2017).

A chi-squared test for independence is a statistical test used to determine whether there is a significant association between two categorical variables. It assesses whether the observed frequencies of the categories within each variable are significantly different from what would be expected if the variables were independent of each other. Similar to a t-test, chi-squared tests were chosen to assess whether there are significant differences in the experiences of chemists from Eng+ vs. Eng1st backgrounds, guided by our research questions. However, unlike a t-test, which requires continuous numerical data given its comparison of means, a chi-squared test allows us to compare categorical data.

Multiple regression is a statistical method used to model the relationship between an outcome variable and one or more independent, predictor variables. Multiple regression offers valuable insights by allowing researchers to analyze the combined effect of multiple predictors on an outcome variable, providing a more comprehensive understanding of complex relationships and facilitating the prediction of outcomes based on several contributing factors. Multiple regression also allows researchers to control for co-variates, which independent variables included in the analysis to control for their potential influence on the relationship between the predictor variables of interest and the outcome variable. Accounting for co-variates helps isolate the unique effect of each predictor variable while accounting for other relevant factors.

We used two types of multiple regression analyses for this study: linear regression and logistic regression. The primary difference between these two types of regression are the outcome variables. In linear regression, the outcome variable is continuous, while in logistic regression,

the outcome variable is categorical. In our study, both types of regression are relevant, as we intend to use these tools to determine the effects of language background (controlling for demographic co-variates). on a combination of continuous and categorical data on (1) chemistry research experiences/productivity, (2) perspectives on and uses of differences language resources, and (3) perspectives and uses on GenAI tools.

7.6 Results and discussion

The following sections describes the findings from preliminary analyses of the questionnaire data.

7.6.1 Research question 1: How does the need to use English in science impact their research experiences and perceived productivity?

Eng+ participants reported encountering more difficulty with various research tasks requiring the use of English ($M = 3.07$, $SD = 1.65$) compared to Eng1st participants ($M = 1.39$, $SD = 0.78$) at levels that were statistically significant, $t(222) = 9.30$, $p < 0.0001$ (Figure 2). Eng+ participants also reported more frequently encountering challenges with English that negatively affected their research productivity and experiences ($M = 2.94$, $SD = 2.02$) compared to Eng1st participants ($M = 1.70$, $SD = 1.46$) at levels that were statistically significant, $t(222) = 5.12$, $p < 0.0001$ (Figure 2).

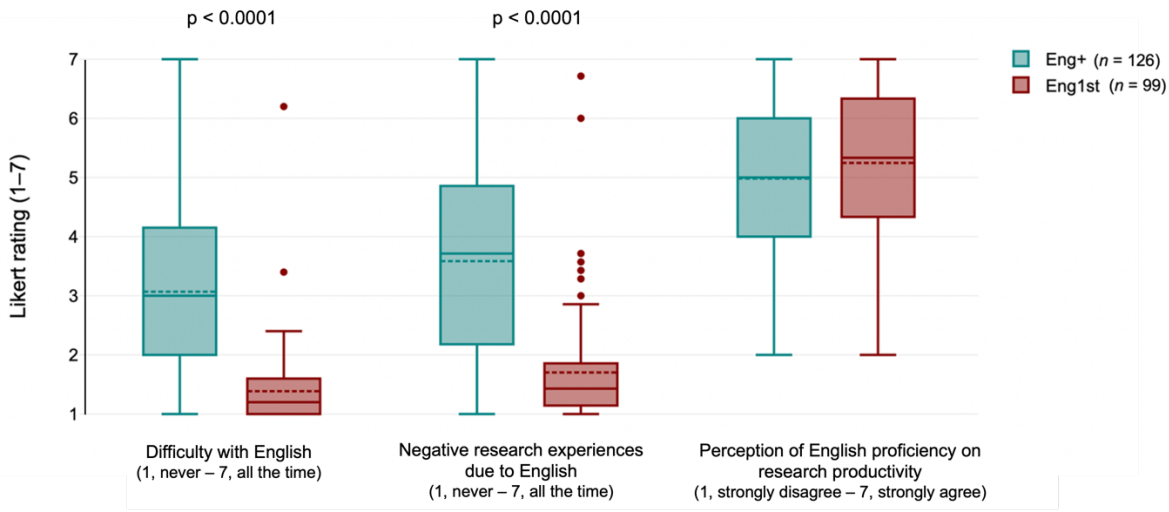


Figure 2: Differences in questionnaire responses between Eng+ chemists (n = 126) and Eng1st chemists (n = 99) to questionnaire sections focused on (1) difficulties with English; (2) challenges with research due to English; and (3) perceptions of English impact on research productivity. Dashed line = mean; Non-dashed line = median.

Section 3 of the questionnaire focused on participants' perceptions of the impact of English on researchers' productivity. We found no statistically significant differences between how Eng+ (M = 4.98, SD = 1.85) and Eng1st participants (M = 5.25, SD = 1.53) responded to this section, $t(222) = 1.14$, $p = 0.254$ (Figure 2).

These data suggest that, first, Eng+ chemistry researchers perceive research tasks that require the use of English to be more difficult than Eng1st chemistry researchers. These findings support existing research that one's language background may shape their experiences in research (Amano et al., 2023; Deng & Flynn, 2023; Ramírez-Castañeda, 2020). Second, the majority of Eng+ and Eng1st chemistry researchers agree that research productivity is positively associated with their English proficiency, including agreement that researchers' language backgrounds contribute to publication success and others' perceptions of their research expertise.

In addition to their ratings, participants were also given the option to elaborate on their perspective in a written response. At the time of writing, we are still analyzing this qualitative data.

7.6.2 Research question 2: What resources and strategies do they use to navigate language barriers during their research?

The goal of the analyses for RQ2 was to identify what resources and supports researchers used to navigate language barriers, including found most/least helpful. Findings from these analyses would allow us to identify what resources/supports researchers find helpful, but less available, allowing our community to develop strategic efforts to support those from diverse language backgrounds.

There was a high positive correlation between reported use and perceived helpfulness (R^2 = 0.83); that is, resources that were used more frequently were also reported to be more helpful. When navigating language barriers during their research, Eng1st participants most reported relying on support from peers and colleagues, their supervisor/mentor, and/or friends and family. These were also reported as the most helpful supports by Eng1st participants.

Translation software (*e.g.*, Google Translate, DeepL) and writing assistant software (*e.g.*, Grammarly) had the largest difference in reported use, with 37% and 32% more Eng+ respondents reporting using these resources compared to Eng1st respondents. Both these resources were rated as more helpful by Eng+ respondents compared to Eng1st respondents at statistically significant levels (translation software, $t(148) = 2.0313$, $p < 0.05$; writing assistant software, $t(126) = 3.4034$, $p < 0.001$). Eng+ respondents reported using GenAI tools 20% more than Eng1st, and Eng+ respondents also perceived GenAI tools to be more helpful ($t(127) = 4.0786$, $p < 0.0001$) (Figure 3).

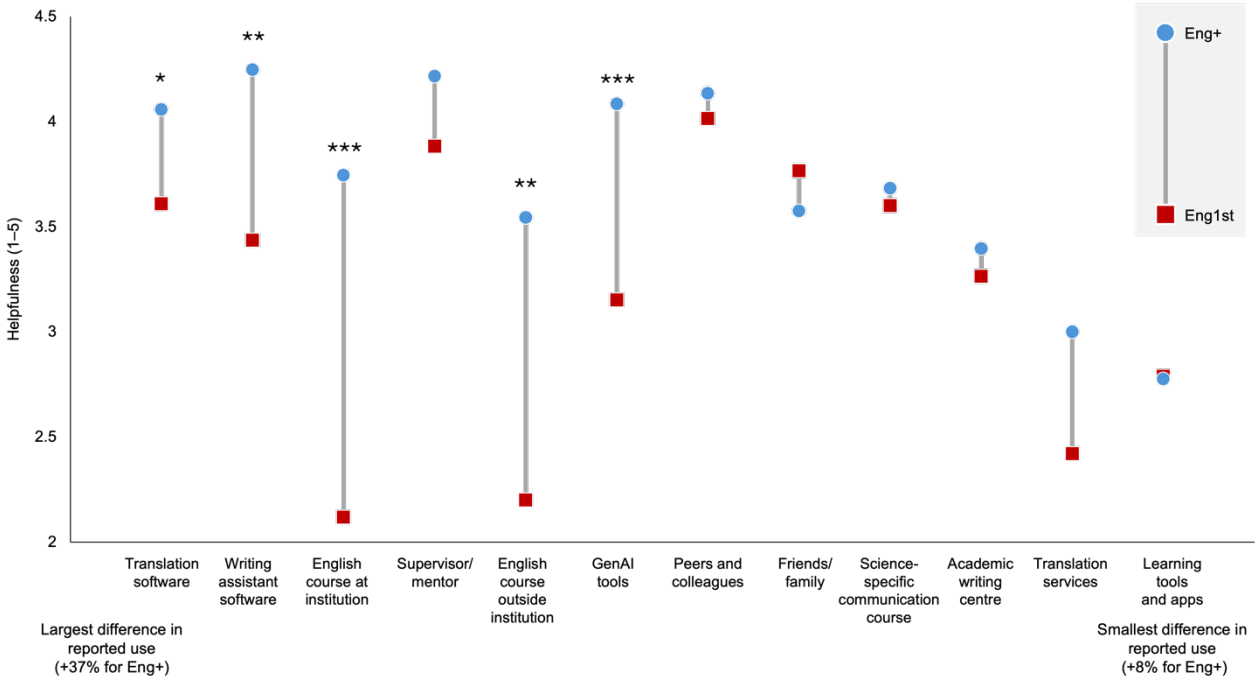


Figure 3: Differences in reported helpfulness of different resources between Eng+ and Eng1st respondents. *$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$

These data suggest that Eng+ respondents view digital resources—such as writing assistant, translation, and GenAI tools—as more helpful compared to Eng1st respondents. One potential reason for this finding is that these digital tools are the most immediately available resource to researchers. For example, a GenAI tool like ChatGPT or a writing assistant like Grammarly often provides an avenue for quicker and more immediate feedback compared peers/colleagues. In this case, one's perception of "helpfulness" may therefore include both the resource's effectiveness in helping researcher's achieve specific goals (*e.g.*, getting feedback on writing), as well as the efficiency or availability of the resource. This hypothesis may be supported by the fact that all of the digital resources reported as being more helpful (at statistically significant levels) for Eng+ respondents were also reportedly used more frequently by Eng+ respondents.

We also found that English courses at and outside respondents' institutions were rated as more helpful by Eng+ respondents compared to Eng1st respondents at statistically significant levels (at institution, $t(76) = 4.9987$, $p < 0.0001$; outside institution, $t(60) = 3.5900$, $p < 0.001$). One

potential reason for this finding is that Eng1st respondents may not rely on or feel the need to attend English language courses to the same degree as those learning English-as-an-additional language.

Peers/colleagues and supervisors/mentors were the resources with the most reported use by all participants overall. However, we did not find any statistically significant differences in terms of helpfulness between the two groups.

Science-specific communication courses were the highest reported "not available" resource/support amongst Eng+ participants. Our previous work found that discipline-specific writing supports (*e.g.*, chemistry communication course) may be a valuable support for both Eng+ and Eng1st researchers in developing an understanding of norms and approaches for research communication (*e.g.*, writing manuscripts, presenting at conferences) (Deng & Flynn, 2023).

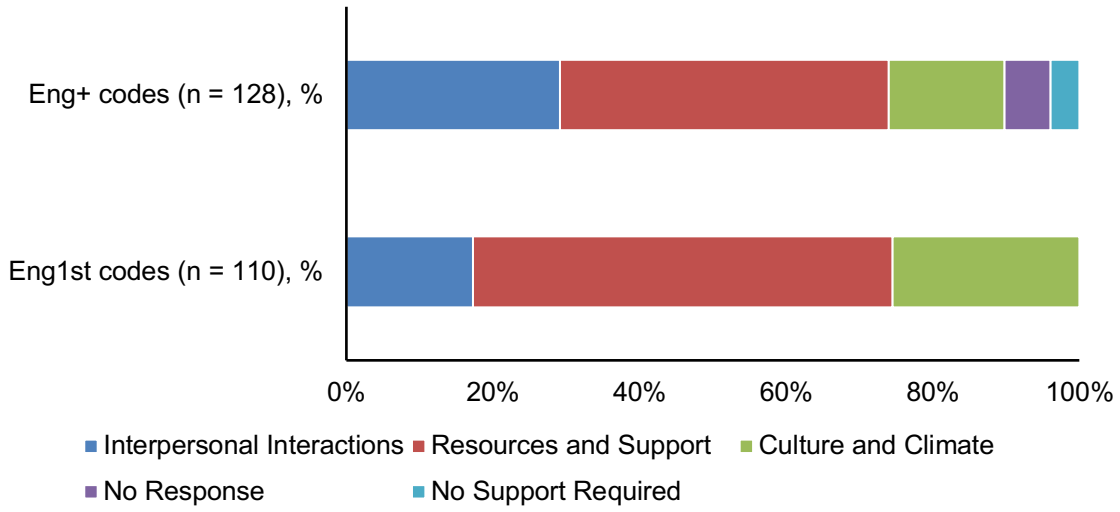


Figure 4: Overview of codes from participants' responses to "How do you think others in thes scientific community could better support scientists who are learning English, if at all? Please try to be as detailed as possible."

Figure 4 shows an overview of participants' responses to the questionnaire item "How do you think others in the scientific community could better support scientists who are learning English,

if at all? Please try to be as detailed as possible." Compared to Eng1st responses, Eng+ responses more frequently discussed the importance of empathetic and patient interpersonal interactions. For example:

> Participant 115 (Eng+): To be more open-minded about their limited knowledge of English and to be inviting when they speak. Or even to find the right words for them. To educate instead of reprimanding them.
>
> Participant 32 (Eng+): Treat them as equals. Give them more time to talk instead of talking over/for them. Just because someone is not fluent in English does not mean they are inferior, less intelligent, or less knowledgeable in science. Also, do not mock or yell at people for having an accent, or for using the wrong words or making grammar mistakes in English.
>
> Participant 10 (Eng+): Use easy vocabulary when talking to these scientists. Also, be patient when people have struggles understanding.

Eng1st responses focused more frequently on the importance of providing resources in the form of writing centres, language and communication courses, and translation/feedback services from journals:

> Participant 31 (Eng1st): I think programs like science-specific language courses and hiring multi-lingual editors of journals is good. For a scientist/researcher on a daily basis, we should try to learn more about other cultures and embrace colleagues who speak other languages. While language barriers certainly cause miscommunications, cultural differences are also a large factor. We need to learn about each other more and view our differences as positives.
>
> Participant 66 (Eng1st): Support networks could be designed, where students learning English could get mentorship both from other English learnings as well as native English-speaking mentors. Other systems like programs designed to practice English would be good too.
>
> Participant 92 (Eng1st) I believe a translation/communication department for each academic institution would be useful. Add professionals who speak multiple languages and can help provide support. Another alternative is to ask research journals to include a translation feature. I am aware that some foreign

> institutions who hire professors will provide a personal secretary/translator for a few years as part of the new professor's welcome package while they get acquainted with their move and new lifestyle. (My knowledge of this is Japanese research institutions.) AI should, in my opinion, be kept as a last resort once all other options for supports are expended.

Interestingly, respondents who believed that no support was required were exclusively Eng+. These respondents argued that the responsibility to learn English was not the responsibility of the scientific community, rather the individual:

> Participant 22 (Eng+): The responsibility is on the scientist, not on the community.
>
> Participant 107 (Eng+): Learning a language is a different matter. I don't think the scientific community needs to invest effort in this. Better translators will handle the task. Above all, what matters is the science. Respect the science, not the language.
>
> Participant 67 (Eng+): I don't think the scientific community should be the ones to support scientists learning English. The only thing they could do is use simple words in paper and avoid the use of high standard English that one needs to translate in order to understand. From experience even native English speaker find the use of fancy English words irritating as it interrupts the flow of your reading every time you try to understand the meaning of 1/4 words.

Together, our analyses provide us insights into both what language resources/supports chemistry researchers currently use and find useful, as well as what additional resources/supports they need from colleagues, supervisors, institutions, and the broader chemistry community.

7.6.3 Research question 3: In what ways are participants using generative artificial intelligence, if at all?

In response to the item "In what ways have you primarily used generative artificial intelligence (*e.g*. ChatGPT) for language support, if at all?", Eng1st respondents more likely to report "None" compared to Eng+ respondents, and a chi-squared test of independence showed that this difference was statistically significant, X^2 (1, N = 225) = 8.29 , p < .005. Eng+ respondents were

more likely use to GenAI for all listed tasks, with the most common use being proofreading (Figure 5).

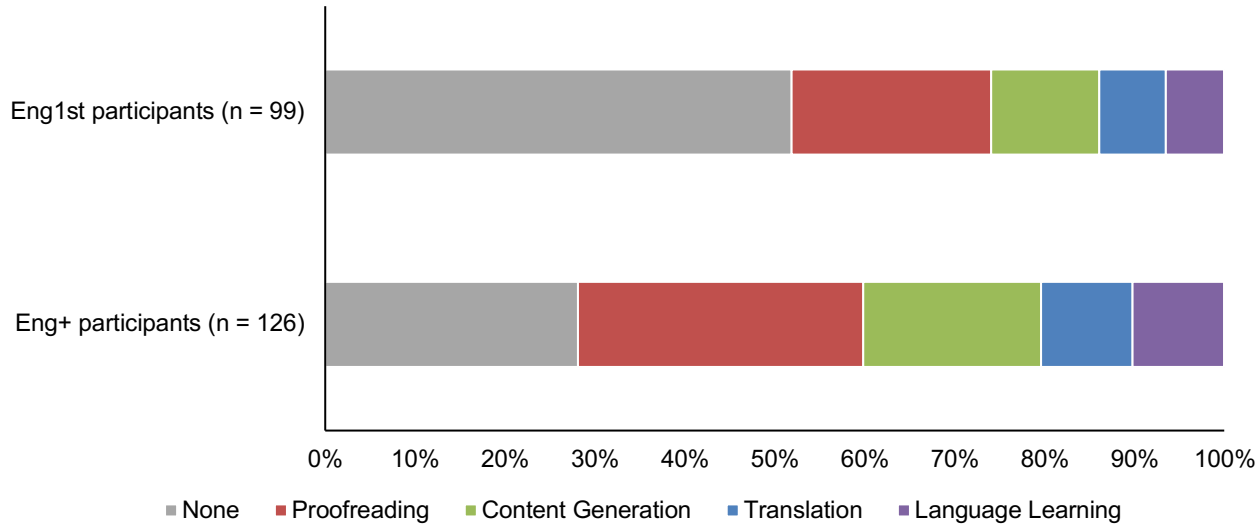


Figure 5: Reported uses of GenAI tools for language support in research by both Eng1st (n = 99) and Eng+ (*n* = 126) participants.

Following this item, respondents were asked the following series of questions:

- Please explain why you have used GenAI for the tasks you chose in the previous question (Q8). If you chose "None", please explain why you have not used GenAI. Please try to be as detailed as possible.
- What benefits or advantages could you see in using GenAI for language support? Please try to be as detailed as possible.
- What limitations or challenges could you see in using GenAI for language support? Please try to be as detailed as possible.

These data were qualitatively analyzed using the process described in the Methods section of this chapter. The following sections will present the overarching themes in these data.

7.6.3.1 Communicating in chemistry

GenAI's most prevalent use across both groups was as a writing support that they integrated into their writing workflow (Figure 6). Respondents primarily reported using GenAI as a proofreading tool. They used tools like ChatGPT to get feedback on writing *after* having already come up with the initial ideas or a draft. Proofreading took in several different forms: (1) grammar, in which respondents sought feedback on grammatical errors or typos; (2) structure, form, or clarity, in which respondents sought more holistic feedback on how to make a passage or paragraph smoother or clearer; and (3) register, in which respondents sought feedback on making their writing sound more formal or academic. Register was more prominent in Eng+ responses, with several respondents specifically talking about using tools like ChatGPT to help them make their professional correspondences (*i.e.*, emails) more professional.

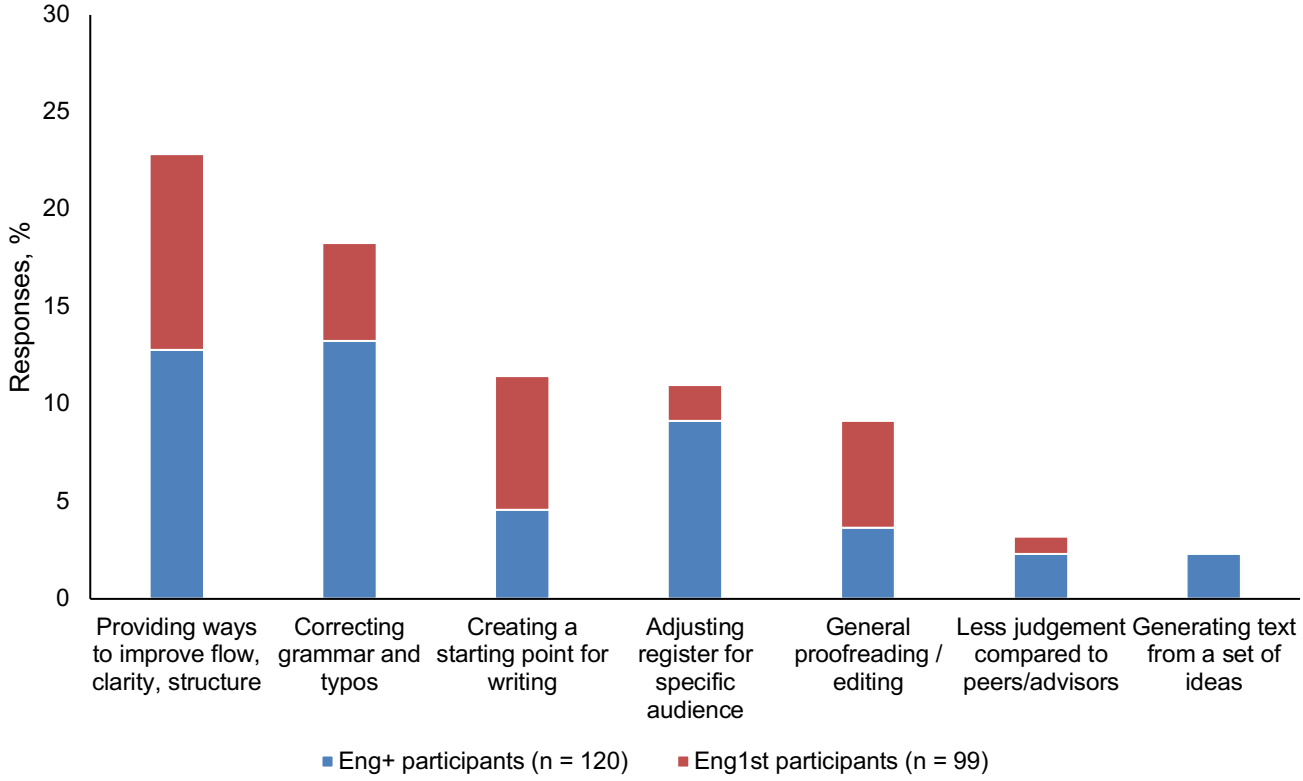


Figure 6: Uses and benefits of GenAI tools for communicating in chemistry reported by participants.

A number of respondents described using GenAI to initiate writing by establishing a starting point (*e.g.*, getting over writer's block). These respondents acknowledged that they would prompt tools like ChatGPT to generate a block of text from an initial set of ideas, using this text as a starting point that they would edit themselves afterwards.

> Participant 60 (Eng1st): My technical writing in English is relatively strong but writing for a more general audience is something I do less frequently and would take me more time. [GenAI can help] proofread documents for appropriate sentence structures; helping to overcome "blank page" intimidation (especially if English is a second language) by providing initial content from which editing/writing can then take place; rewriting technical content into content for a more general audience.

Several participants discussed a preference for GenAI tools for writing feedback over colleagues/advisors because they viewed these online tools as less judgmental:

> Participant 56 (Eng1st): AI doesn’t judge you for not understanding. Profs have intrinsic bias.

> Participant 69 (Eng+): Sometimes in academia; you might have challenges, fears and achievements that others (out of the scientific community) might not full[y] understand your feelings and you might not feel intimate enough with colleagues to talk about the topic. In this regards [sic], AI helps by listening!

7.6.3.2 Learning in chemistry

Using GenAI tools to support learning in chemistry focused on how GenAI tools could be used to re-phrase or translate scientific text (Figure 7).

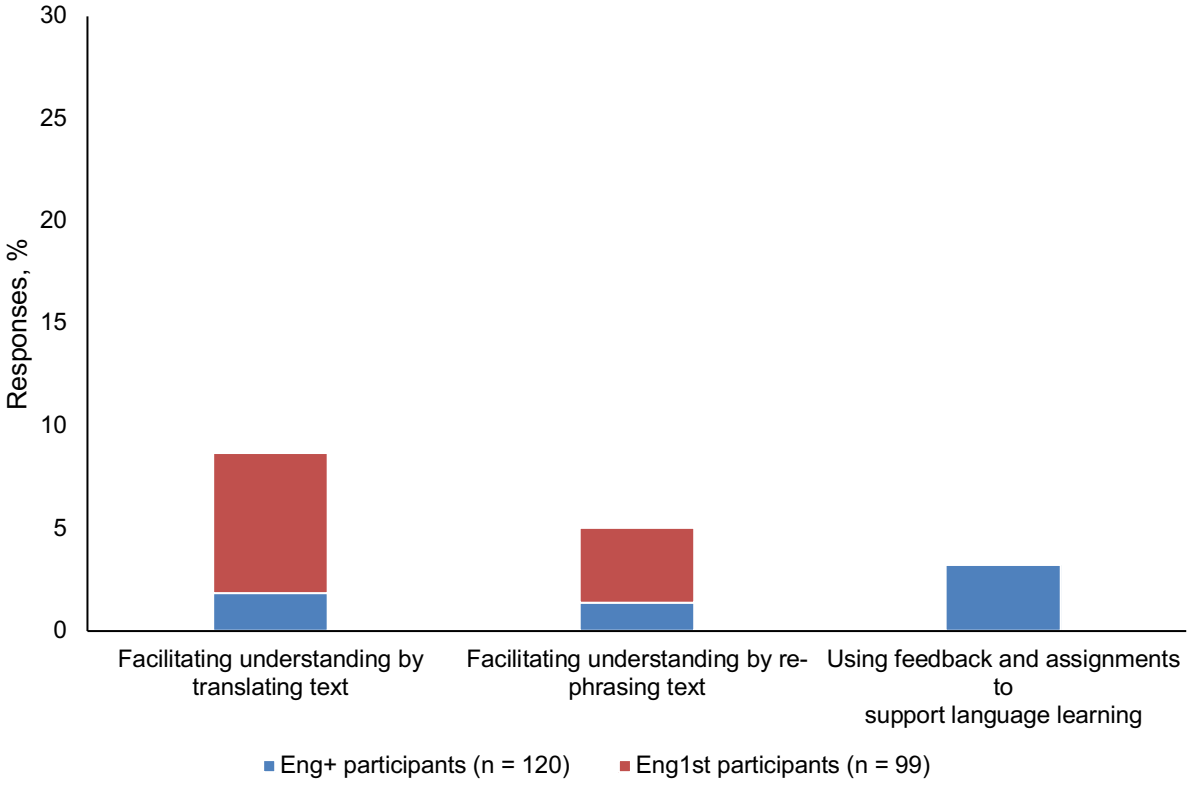


Figure 7: Uses and benefits of GenAI tools for learning in chemistry reported by participants.

An example of re-phrasing text included inputting text into ChatGPT from a scientific paper and prompting ChatGPT to re-phrase the text into simpler terms:

> Participant 51 (Eng+): I got [sic] benefit by using GenAI while doing my assignments and reading journal because sometimes I didn't understand the sentences properly and use GenAI to clarify it more simple way.
>
> Participant 97 (Eng1st): Simplifying convoluted research papers - some papers are written needlessly complex.
>
> Participant 42 (Eng1st): Summarizing complex text into simple messaging for better understanding, or drafting manuscripts or reports from large amounts of data.

Others who used GenAI tools to translate scientific text effectively used these tools as a translation service, helping them translate texts into other languages to facilitate understanding:

> Participant 5 (Eng1st): I have used this tool to translate patents from other languages. Translating patents or manuscripts from other languages and gathering background on a topic.
>
> Participant 69 (Eng1st): I am not able to speak or write in Spanish, yet some of my dissertation data is in Spanish. So I often use GenAI to translate these sections so I can analyze my data in English, while honoring my participants' choice of language. GenAI could eventually help Eng+ scientists with making sure that their writing is grammatically correct without requiring review by a human. Scientific jargon is dense and often inaccessible, and I am not confident that GenAI will understand the meaning and context of some terminology.
>
> Participant 20 (Eng+): I need AI helps me to translate some strange words and sentences that I feel confused.

7.6.3.3 Doing chemistry research

The most commonly reported benefit or use from using GenAI tools when doing chemistry research was as a way to support research productivity (Figure 8).

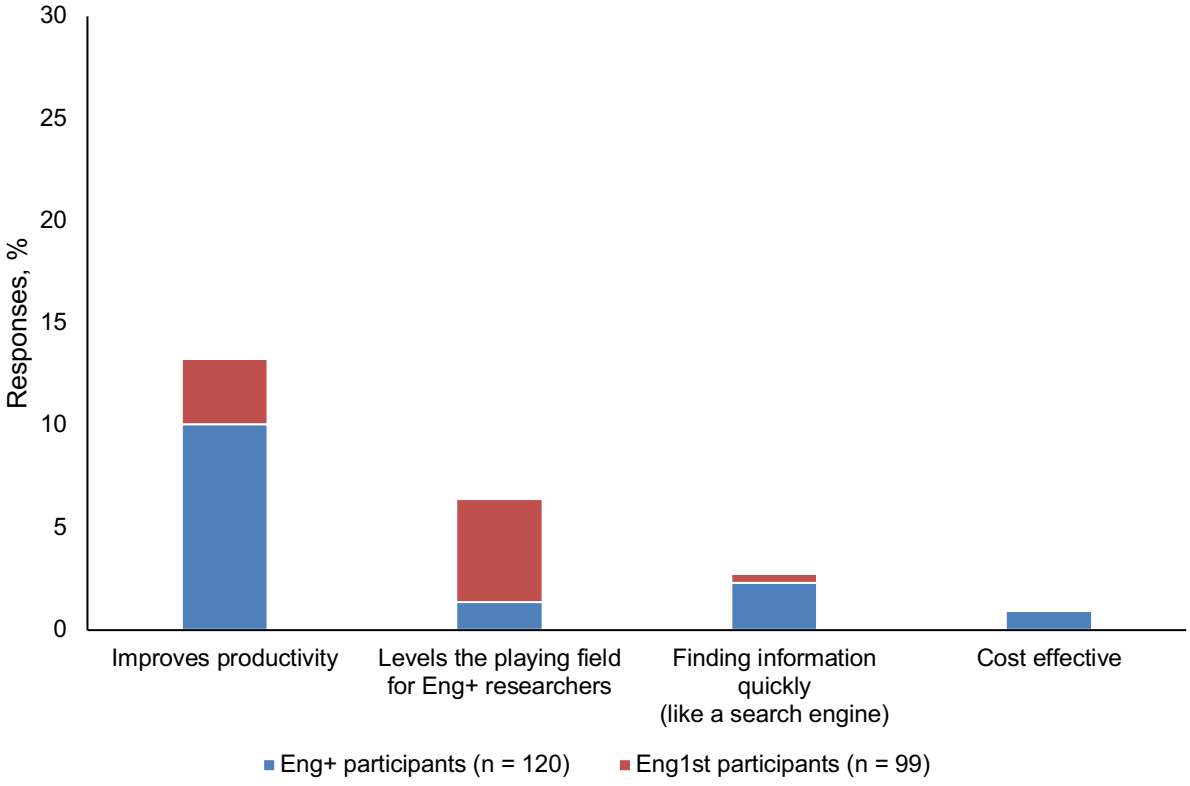


Figure 8: Uses and benefits of GenAI tools for doing chemistry research reported by participants.

Many of these responses commented on how by helping with writing, the researchers could spend more time doing other research tasks. Our previous work also demonstrated that challenges in writing were associated with research delays amongst Eng+ chemistry trainees (Deng & Flynn, 2023)—similar sentiments were expressed in this study:

> Participant 105 (Eng+): It's much easier to learn through GenAI because I think I get the full tasks done and in a very simpler way for me to understand. I can only think of benefits for now sure[,] it saves me alot of time. The only challenge I think I might face is maybe having issues with the software security

> Participant 48 (Eng+): It saves time. And I am sure that the language is more formal and understandable. It is very beneficial for research being cost effective. But as a way to learn language

> it might have some cons. Your language level might not develop anymore.
>
> Participant 111 (Eng+): Using GenAI increases the productivity and efficacy of my research. Additionally, it improves the style and tonality of my writing and helps avoid common mistakes. GenAI can check for mistakes and improve the style of one's writing. It helps paraphrase sentences and find synonyms for words. It increases the clarity of paragraphs. GenAI can hallucinate and is not to be used as a reliable source for generating novel content.
>
> Participant 107 (Eng+): I typically use ChatGPT to refine my content, correcting grammar mistakes, eliminating redundant word usage, and improving overall clarity. I don't usually employ it to interpret data. I have not used any other tools. Using ChatGPT saves a significant amount of time, and I have more confidence in the text it generates compared to the assistance I receive from my colleagues. Sometimes, the language used by ChatGPT is not appropriate for the research article. It can be overly dramatic or use alternative words for scientific phenomena, thereby completely changing the meaning.
>
> Participant 70 (Eng1st): Even though I am a native English speaker, I sometimes feel like I could use some help from an AI to fix my errors. I see this helping non-native English speakers tremendously. All of my colleagues are non-native English speakers and I see them use ChatGPT all the time.

A few participants (all Eng1st) also discussed how GenAI tools could be used to support with coding:

> Participant 21 (Eng1st): ChatGPT is quite effective at translating pseudo code into working code in many different languages. It can also do a reasonable job of correcting certain types of

correcting simple code and explaining why. I could see it as being useful for generating more readable and better flowing text in one's second or third language. Issues with mistranslation or improper connotations being used, impeding a more complete mastery of the language due to an overreliance on the tool

7.6.3.4 Concerns and reasons for not using GenAI

Despite reported uses and opportunities for research presented by GenAI tools, participants also discussed many reasons for being wary of or not using these tools, including awareness of these tools' limitations and ethical concerns (Figure 9).

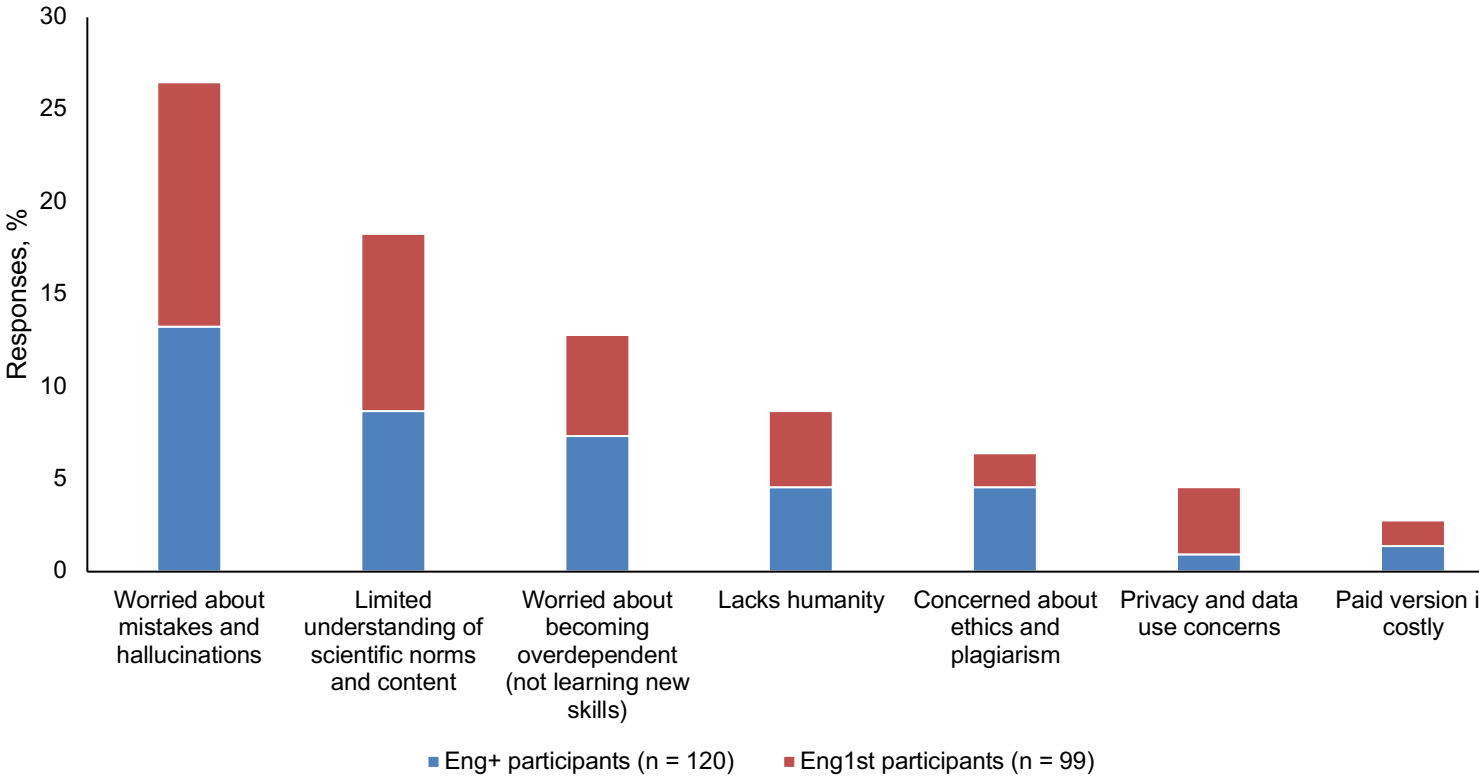


Figure 9: Limitations and concerns about GenAI tools for chemistry research reported by participants.

Across all ideas expressed in participants' responses, the two most prevalent ideas about GenAI expressed by participants were concerns about ethics and plagiarism and GenAI tools limited understanding of scientific norms and content. For example, participants said:

> Participant 31 (Eng+): Scientific writing is very delicate and a word change can alter the meaning completely. I do not trust AI to understand and preserve the meaning. In my line of work I do not need to waste time writing corporate fluff, so I have no need for AI. Only for non-scientific correspondence, such as invites etc., where more flowery language might be needed. For scientific publications, spellcheck and Word's suggestions should be enough and if not, a scientist proficient in English and familiar with the research should edit the text. AI tools don't understand science and can't be trusted to edit scientific texts.

> Participant 110 (Eng+): I don't trust GenAI tools to create reliable text. Also, I believe there is a big difference when you write your document rather than having it done for you by an AI tool. It could help with proofreading a text but not further than that. Cannot fact check in a reliable manner. Specially when it comes to advanced research topics, the information is not readily available even to an AI tool.

GenAI tools already brought forth complex discussions and debates within chemistry education and research. Concerns around academic integrity in research have led to some journals banning their use, while others are still working to develop regulations and policies (Elsen-Rooney, 2023; Nature, 2023b). In education, there is a growing body of literature that argues for equipping learners with the skills and knowledge to use GenAI tools safely and ethically (Holmes, 2023; Wiley, 2023). Data from our questionnaire suggests that researchers from both Eng+ and Eng1st backgrounds may already be aware and mindful of GenAI's limitations, including their propensity to generate false information and the risks of academic plagiarism.

Other reasons for not using GenAI centred on participants not *needing* to rely on such tools for language support (Figure 10).

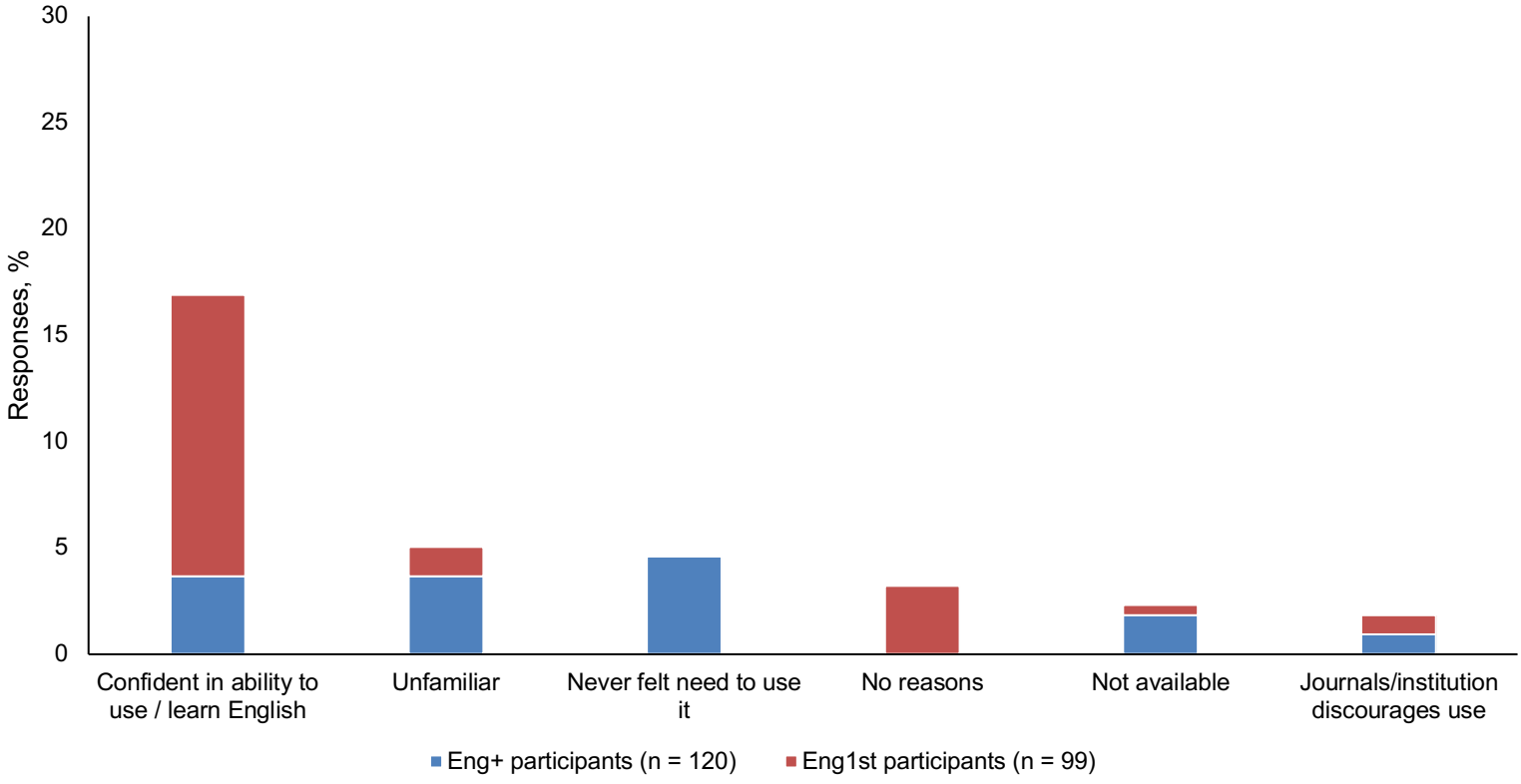


Figure 10: Reasons for not using GenAI tools in chemistry research reported by participants.

For example, many Eng1st participants simply felt they were good enough at writing that they did not feel the need to use GenAI tools to support their writing. Others described feelings about how GenAI lacked a "human touch" when writing and that human writing is often more creative and flexible.

There were also instances of Eng+ participants who mentioned not *wanting* to use GenAI tools for language support. These participants' responses focused on concerns about becoming overdependent on GenAI, expressing a desire to develop their language skills even if it is an additional challenge.

> Participant 12 (Eng+): I don't use this tool because I would like to improve my capabilities in English than use the tool. I know it could be quickly but I know sometimes it can make mistakes. Is a quickly method to improve a text in another languages, also you can choose the intention of the document and specify the

> destinatary [sic]. It could have mistakes, is necessary to be critic and to be capable of evaluate the text generated.
>
> Participant 48 (Eng+): It saves time. And i am sure that the language is more formal and understandable. It is very beneficial for research being cost effective. But as a way to learn language it might have some cons. Your language level might not develop anymore.

This particular finding demonstrates Eng+ chemistry trainees growth mindset in navigating English language barriers in research, which reinforces a finding from our previous work (Deng & Flynn, 2023).

7.7 Conclusions, implications, and future work

We developed and administered a questionnaire to collect and analyze quantitative and qualitative data to study the experiences of chemistry researchers from diverse language backgrounds Though our data analysis is still to be completed, there are a few preliminary findings from this work that present implications for both chemistry research and education.

For research question 1, we found that Eng+ reported higher levels of difficulty with research tasks in English compared to Eng1st researchers at statistically significant levels. Both Eng+ and Eng1st chemistry researchers viewed proficiency in English to be associated with research productivity.

For research question 2, digital tools, such as translation software, writing assistant software, and GenAI tools, were perceived as more helpful by Eng+ respondents compared Eng1st at statistically significant levels. These findings suggest that digital tools, including GenAI tools could be advantageous for supporting linguistic equity in research (Berdejo-Espinola & Amano, 2023). Both groups relied greatly on their peers and colleagues and supervisors or mentors for language support. Given participants' reliance on their immediate professional circles, the nature of mentorship practices is likely to have a high impact. Research evidence and resources outlining

specific strategies that support inclusive and evidence-based mentorship in science are growing (Barrette-Ng et al., 2019; Ceyhan & Tillotson, 2020; Deng & Flynn, 2023; Lopez & Duran, 2021; National Academies of Science Engineering and Medicine, 2019; Paglis et al., 2006; Sarabipour et al., 2023; Schmid & Wiles, 2022); future work could investigate the impact and efficacy of different mentorship strategies on researchers from diverse language backgrounds, and/or the effects of targeted interventions to equip researchers with mentorship skills (Barette-Ng et al., 2024).

For research question 3, we found that both Eng+ and Eng1st participants primarily used GenAI tools as a writing support. Most frequently, participants used GenAI to edit and provide feedback on their writing, including in scientific communications (*e.g.*, manuscripts, abstracts) and professional communications (*e.g.*, email). The use of GenAI as a writing support was often associated with views that GenAI tools supported their productivity, as participants felt like they could spend less time on writing and more time engaging in other research tasks. We also found that both groups held concerns about the limitations and ethical considerations of GenAI tools. In particular, participants shared concerns about ethics and plagiarism and GenAI tools limited understanding of scientific norms and content, leading them to be wary of using GenAI tools for learning information or generating content. Lastly, some Eng+ participants expressed not wanting to use GenAI tools because they did want to become over reliant on these tools for support. These participants expressed a desire to learn English, despite it presenting unique challenges and barriers for them, indicative of a growth mindset (Deng & Flynn, 2023). Our findings add research evidence to the existing body of editorials and perspectives on the use of GenAI tools for scientific writing, which have often focused exclusively on the ethical concerns and challenges posed by the emergence of these tools (Nature, 2023a; Rahman & Watanobe, 2023; Stokel-Walker, 2023; van Dis et al., 2023). As well, our findings add to a growing library of research on how researchers use GenAI tools, with some studies suggesting primarily limited or non-research

use (Owens, 2023) and others relying on these tools as an essential part of their research, such as in designing metal-organic frameworks (Zheng et al., 2023).

We explored how language shapes research experiences *between* those from Eng1st and Eng+ backgrounds. Other work has exclusively focused on Eng+ researchers; without a comparison between different groups, we would not know to what extent experiences differ (if at all) between those from different language backgrounds. Previous research has been conducted in the biological sciences in Europe, Asia, and South America, and our study expands on this previous work by examining the role of language in scientific research in North America. This context is especially important, as Canada and the US are global leaders in scientific research and English-dominant regions of the world with many citizens and scientists from diverse language backgrounds. This work also examined via a questionnaire how researchers from diverse language backgrounds use and perceive different language resources available to them. These findings will be essential in the strategic design and evaluation of resources to support both Eng1st and Eng+ researchers, aligned with their priorities and perspectives. Lastly, this work is the first to examine perspectives on GenAI tools amongst those from diverse language backgrounds. The potential for these tools to support linguistic equity in science has been discussed previously (Berdejo-Espinola & Amano, 2023), but prior to this work, we knew little from research on how researchers were using GenAI tools or their perspectives on these tools for language support in research. The implications from this finding have the potential to guide ongoing discussions about how we as a scientific community might use GenAI tools both safely and in ways that support equity.

Future steps in data analysis include the planned regression analyses. Such analyses will provide us with deeper insights into the how one's English language proficiency/history impacts their perspectives and experiences in chemistry, beyond dichotomous Eng1st and Eng+ labels. Regression analyses will also allow us to control for and evaluate the influences of other demographic data collected from the questionnaire, such as race/ethnicity, career stage, and

country of work (*e.g.,* how might career stage interact with language background to shape chemistry researchers' experiences?). Completion of the qualitative analyses in this study will also provide us with more nuanced insights into participants' experiences and views on the role of language in chemistry research.

Our findings thus both complement and expand on existing research evidence on the experiences of researchers from diverse language backgrounds. This work is the first to identify how emerging technologies like GenAI are viewed by researchers from different language backgrounds, including how they could be used as tools to enhance equity and productivity for Eng+ researchers. Organizations and institutions around the world are grappling with GenAI tools, with many of these conversations centred on concerns about academic integrity and plagiarism. Our research findings add to these important conversations, demonstrating that (1) these tools may serve as effective and accessible writing supports for both Eng+ and Eng1st researchers; and (2) many researchers are aware of the limitations of these tools for learning and research, and are mindful of these limitations when choosing to (not) use these tools.

7.8 Limitations

Questionnaires rely on participants to self-report data, and this type of data may be subject to various biases, such as social desirability bias (Edwards, 1957). Participants may have described their experiences or views differently compared to their actual experience or views due to a desire to be viewed a certain way. To mitigate social desirability bias, we stated in the consent form for the questionnaire that "The information you share will remain strictly confidential. We will NOT share any of this information with anyone else at any time. We will ensure anonymity by referring to me using a pseudonym in all communications, including any quotes. Participating in this study will have no impact on your academic or career standing."

Data was collected via convenience sampling. Participants voluntary participated through a publicly available online questionnaire. This data collection approach is subject to self-selection

bias; that is, participants may have unique characteristics that led them to participate in the study, thus creating a sample that is less representative of the broader population (Elston, 2021). Though we tailored our data recruitment strategies for those who fit the study's inclusion criteria (i.e., chemistry researchers from diverse language backgrounds), some participants may not have fit the inclusion criteria.

7.9 References for Chapter 7

Abedi, J. (2002). Assessment and accommodations of English language: Issues, concerns, and recommendations. *Journal of School Improvement*, *3*(1), 83–89.

Abedi, J. (2015). Language Issues in Item Development. *Handbook of Test Development*, *Civ*, 355–373. https://doi.org/10.4324/9780203874776.ch17

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for english language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*(1), 1–28. https://doi.org/10.3102/00346543074001001

Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*, *14*(3), 219–234. https://doi.org/10.1207/S15324818AME1403

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining Effectiveness and Validity of Accommodations for English Language Learners in Mathematics: An Evidence-Based Computer Accommodation Decision System. *Educational Measurement: Issues and Practice*, *39*(4), 41–52. https://doi.org/10.1111/emip.12328

Afitska, O., & Heaton, T. J. (2019a). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, *103*(6), 1396–1422. https://doi.org/10.1002/sce.21545

Afitska, O., & Heaton, T. J. (2019b). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, 1396–1422. https://doi.org/10.1002/sce.21545

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *Journal of Chemical Education*, *100*(4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00323

Allensworth, E. (2005). Graduation and Dropout Trends in Chicago: A Look at Cohorts of Students from 1991 through 2004. In *Consortium on Chicago School Research* (Issue January). papers://e97baf73-a10d-438d-985b-bee59744e11f/Paper/p939

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages Are Still a Major Barrier to Global Science. *PLoS Biology*, *14*(12), e2000933. https://doi.org/10.1371/journal.pbio.2000933

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., David González-Trujillo, J., Montaño-3 Centellas, F., Paudel, K., White, R., & Veríssimo, D. (2023). The cost of being a non-native English speaker in science. *PLoS ONE*, *21*(7), 1–27.

Amano, T., Rios Rojas, C., Boum II, Y., Calvo, M., & Misra, B. B. (2021). Ten tips for overcoming language barriers in science. *Nature Human Behaviour*. https://www.nature.com/articles/s41562-021-01137-1

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Ammon, U. (Ed.). (2011). *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110869484

Angulo, E., Diagne, C., Ballesteros-Mejia, L., Adamjy, T., Ahmed, D. A., Akulov, E., Banerjee, A. K., Capinha, C., Dia, C. A. K. M., Dobigny, G., Duboscq-Carra, V. G., Golivets, M., Haubrock, P. J., Heringer, G., Kirichenko, N., Kourantidou, M., Liu, C., Nuñez, M. A., Renault, D., … Courchamp, F. (2021). Non-English languages enrich scientific knowledge:

The example of economic costs of biological invasions. *Science of the Total Environment*, *775*, 144441. https://doi.org/10.1016/j.scitotenv.2020.144441

Barbera, J., & VandenPlas, J. R. (2011). All assessment materials are not created equal: The myths about instrument development, validity, and reliability. *ACS Symposium Series*, *1074*, 177–193. https://doi.org/10.1021/bk-2011-1074.ch011

Barette-Ng, I., Bourgonje, C., Cheng, T., DiLabio, G., Flynn, A., Hellemans, K., Hinch, I., Kerr, J., McMunn, L., McPhedran, S., Pillon, A., Thompson, A., & Wuetherick, B. (2024). *Empowering and equipping HQP in STEM*. https://www.flynnresearchgroup.com/hqp

Barrette-Ng, I. H., Nowell, L., Anderson, S. J., Arcellana-Panlilio, M., Brown, B., Chalhoub, S., Clancy, T. L., Desjardine, P., Dorland, A., Dyjur, P., Mueller, K., Reid, L., Squance, R. T., Towers, J., & Wilcox, G. (2019). *The Mentorship Guide for Teaching and Learning*.

Bazeley, P., & Jackson, K. (2013). Qualitative Data Analysis with NVivo. In *Journal of Education for Teaching* (Vol. 40, Issue 2). Sage Publications Inc. https://doi.org/10.1080/02607476.2013.866724

Berdejo-Espinola, V., & Amano, T. (2023). AI tools can improve equity in science. *Science*, *379*(6636), 991. https://doi.org/10.1126/science.adg9714

Botten, E. L. (2023). *How to use ChatGPT as a learning tool*. American Psychological Association. https://doi.org/10.12968/bjon.2012.21.6.361

Buxton, C. A., Allexsaht-Snider, M., Aghasaleh, R., Kayumova, S., Kim, S.-H., Choi, Y.-J., & Cohen, A. (2014). Potential Benefits of Bilingual Constructed Response Science Assessments for Understanding Bilingual Learners' Emergent Use of Language of Scientific Investigation Practices. *Double Helix*, *2*(September 2016), 1–21.

Buxton, C. A., Salinas, A., Mahotiere, M., Lee, O., & Secada, W. (2015). Fourth-grade emergent bilingual learners' scientific reasoning complexity, controlled experiment practices, and content knowledge when discussing school, home, and play contexts. *Teachers College Record*, *117*(2).

Ceyhan, G., & Tillotson, J. (2020). *Mentoring Structures and the Types of Support Provided to Early-Year Undergraduate Researchers*. https://doi.org/10.1187/cbe.19-09-0183

Cheng, Y., Dharwadkar, S., Sheridan, V., Rojas, C. R., Bosch Grau, M., & Gordin, M. (2019). Science's language barrier. *Nature*, *570*, 265.

Chin, W. Y. (2010). Linguistic profiling in education: how accent bias denis equal educational opportunites to students of color. In *The Scholar* (Vol. 12, Issue 355, pp. 356–384).

Clark, T. M. (2023). Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.3c00027

Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 443–507). American Council on Education.

Crossman, E., Lu, Y., & Hou, F. (2022). International students as a source of labour supply: Engagement in the labour market after graduation. *Economic and Social Reports*, *2*(3), 1–5.

Curtis, S., & Millar, R. (1988). Language and Conceptual Understanding in Science: A comparison of English and Asian language speaking children. *Research in Science & Technological Education*, *6*(1), 61–77. https://doi.org/10.1080/0263514880060106

Deng, J. M., & Flynn, A. B. (2023). "I Am Working 24/7, but I Can't Translate That to You": The Barriers, Strategies, and Needed Supports Reported by Chemistry Trainees from English-as-an-Additional Language Backgrounds. *Journal of Chemical Education*, *100*(4), 1523–1536. https://doi.org/10.1021/acs.jchemed.2c01063

Deng, J. M., Rahmani, M., & Flynn, A. B. (2022). The role of language in students' justifications of scientific phenomena. *International Journal of Science Education*, *44*(13), 2131–2151. https://doi.org/10.1080/09500693.2022.2114299

Deng, J. M., Streja, N., & Flynn, A. B. (2021). Response Process Validity Evidence in Chemistry

Education Research. *Journal of Chemical Education*, *98*(12), 3656–3666. https://doi.org/10.1021/acs.jchemed.1c00749

Dietrich, S., & Hernandez, E. (2022). *Nearly 68 Million People Spoke a Language Other Than English at Home in 2019*. Census Bureau. https://www.census.gov/Library/Stories/2022/12/Languages-We-Speak-in-United-States.Html

Diversify Chemistry. (2024). *Diversify Chemistry*. https://diversifychemistry.com/

dos Santos, R. P. (2023). Enhancing Chemistry Learning with ChatGPT and Bing Chat as Agents-to-Think-With: A Comparative Case Study. *ArXiv*, 1–27.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. The Dryden Press.

Elsen-Rooney, M. (2023). *NYC education department blocks ChatGPT on school devices, networks*. ChalkBeat. https://ny.chalkbeat.org/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence

Elsherief, H., & Masson, M. (2020). *Every Teacher is a Language Teacher* (Vol. 1). University of Ottawa Second Language Education Cohort (cL2c). https://doi.org/http://dx.doi.org/10.20381/an3d-6k16

Elston, D. M. (2021). Participation bias, self-selection bias, and response bias. *Journal of the American Academy of Dermatology*, 6–7. https://doi.org/10.1016/j.jaad.2021.06.025

Emenike, M. E., & Emenike, B. U. (2023). Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. *Journal of Chemical Education*, *100*(4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00063

Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating Academic Answers Generated Using ChatGPT. *Journal of Chemical Education*, 1–4. https://doi.org/10.1021/acs.jchemed.3c00087

Field, A. (2014). *Discovering Statistics Using IPM SPSS Statistics* (4th ed.). Sage Publications Ltd.

Foote, J. A. (2010). *Teaching with an Accent*.

Gardner, D. E., & Giordano, A. N. (2023). The Challenges and Value of Undergraduate Oral Exams in the Physical Chemistry Classroom: A Useful Tool in the Assessment Toolbox. *Journal of Chemical Education*, *3*, 1705–1709. https://doi.org/10.1021/acs.jchemed.3c00011

González-Howard, M., & McNeill, K. L. (2016). Learning in a community of practice: Factors impacting english-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, *53*(4), 527–553. https://doi.org/10.1002/tea.21310

Government of Canada. (2022). *While English and French are still the main languages spoken in Canada, the country's linguistic diversity continues to grow*. The Daily. https://www150.statcan.gc.ca/n1/daily-quotidien/220817/dq220817a-eng.htm?utm_source=cst&utm_medium=smo&utm_campaign=statcan-2021census-diss-language-en

Higher Education Strategy Associates. (2020). *The state of post-secondary education in Canada*.

Hofstra, B., Kulkarni, V. V., Galvez, S. M. N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Holmes, W. (2023). *Fair Comment: This was not written by ChatGPT*. Commonwe. https://www.col.org/news/fair-comment-this-was-not-written-by-chatgpt/

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments?-A generalizability theory approach. *Assessing Writing*, *13*(3), 201–218. https://doi.org/10.1016/j.asw.2008.10.002

Humphry, T., & Fuller, A. L. (2023). Potential ChatGPT Use in Undergraduate Chemistry

Laboratories. *Journal of Chemical Education*, 1434–1436. https://doi.org/10.1021/acs.jchemed.3c00006

Institute of International Education. (2020). *Project Atlas: A Quick Look at Global Mobility Trends*. https://iie.widen.net/s/g2bqxwkwqv/project-atlas-infographics-2020

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12001

Kember, D., Ha, T. S., Lam, B. H., Lee, A., Ng, S., Yan, L., & Yum, J. C. K. (1997). The diverse role of the critical friend in supporting educational action research projects. *Educational Action Research*, *5*(3), 463–481. https://doi.org/10.1080/09650799700200036

Kirk, T. (2023). *ChatGPT (We need to talk)*. University of Cambridge. https://www.cam.ac.uk/stories/ChatGPT-and-education

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: The cornerstone of modern statistics. *Korean Journal of Anesthesiology*, *70*(2), 144–156. https://doi.org/10.4097/kjae.2017.70.2.144

LaCosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-belonging intervention improves STEM outcomes for students who speak English as a second language. *Science Advances*, *6*(40), 1–11. https://doi.org/10.1126/sciadv.abb6543

Lee, E. N. (2018). *Breaking the language barrier: Equitable assessment in general chemistry (doctoral dissertation)* [University of Nevada, Las Vegas]. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2018-58618-069&lang=de&site=ehost-live

Lee, E. N., Nealy, S., & Cruz, L. (2023). Navigating the interlanguage space: Chinese international students' perceptions of a virtual chemistry laboratory course. *Chemistry Education Research and Practice*. https://doi.org/10.1039/d2rp00145d

Lee, E. N., & Orgill, M. (2021). Toward Equitable Assessment of English Language Learners in General Chemistry: Identifying Supportive Features in Assessment Items. *Journal of*

Chemical Education. https://doi.org/10.1021/acs.jchemed.1c00370

Lee, E. N., Orgill, M., & Kardash, C. (2020). Supporting English Language Learners in College Science Classroom: Insights from Chemistry Students. *Multicultural Education*, *27*(Spring/Summer), 25–32.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-Language backgrounds. *Educational Researcher*, *27*(4), 12–21. https://doi.org/10.3102/0013189X027004012

Lee, O., Quinn, H., & Valdés, G. (2013). Science and Language for English Language Learners in Relation to Next Generation Science Standards and with Implications for Common Core State Standards for English Language Arts and Mathematics. *Educational Researcher*, *42*(4), 223–233. https://doi.org/10.3102/0013189X13480524

Lemke, J. L. (1991). Talking science: language, learning, and values. In *Choice Reviews Online* (Vol. 28, Issue 09). https://doi.org/10.5860/choice.28-5211

Lindsey, P., & Crusan, D. (2011). How Faculty Attitudes and Expectations Toward Student Nationality Affect Writing Assessment. *Across the Disciplines*, *8*(4), 1–19. https://doi.org/10.37514/atd-j.2011.8.4.23

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, *23*(2), 163–198. https://doi.org/10.1017/S0047404500017826

Lopez, P. "., & Duran, A. (2021). *The Role of Mentoring on the Retention of Women From Diverse Backgrounds in STEM*. https://doi.org/10.4018/978-1-7998-4745-8.ch010

Lyon, E. G., Bunch, G. C., & Shaw, J. M. (2012). Navigating the language demands of an inquiry-based science performance assessment: Classroom challenges and opportunities for English learners. *Science Education*, *96*(4), 631–651. https://doi.org/10.1002/sce.21008

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education*, *94*(6), 937–962. https://doi.org/10.1002/sce.20408

Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., & Shevlin, M. (2019). Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. *International Journal of Qualitative Methods*, *18*, 1–11. https://doi.org/10.1177/1609406919878341

McNeil, L. (2005). Faking equity: High-stakes testing and the education of Latino youth. In A. Valanzeula (Ed.), *Leaving children behind: How "Texas style" accountability fails Latino youth* (pp. 57–112). University of New York Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 741–749.

Milnes, T., & Cheng, L. (2008). Teachers' Assessment of ESL Students in Mainstream Classes: Challenges, Strategies, and Decision-Making. *TESL Canada Journal*, *26*(1), 49. https://doi.org/10.18806/tesl.v26i1.129

National Academies of Science Engineering and Medicine. (2018). *Graduate STEM Education for the 21st Century* (A. Leshner & L. Scherer (Eds.)). National Academies Press. https://doi.org/10.17226/25038

National Academies of Science Engineering and Medicine. (2019). *The Science of Effective Mentorship in STEMM* (A. Byars-Winston & M. L. Dahlberg (Eds.)). National Academies Press.

National Science Board. (2021). Publications Output: U.S. Trends and International Comparisons. In *Science and Engineering Indicators 2022*.

National Science Foundation. (2020). *2019 Doctorate Recipients from U.S. Universities: Vol. NSF 21-308* (Issue December 2020). http://www.nsf.gov/statistics/2016/nsf16300/digest/nsf16300.pdf

Nature. (2023a). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, *613*, 612.

Nature. (2023b). Why Nature will not allow the use of generative AI in images and video. *Nature*, *618*(7964), 214. https://doi.org/10.1038/d41586-023-01546-4

Owens, B. (2023). How Nature readers are using ChatGPT. *Nature*, *615*(20).

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451–476. https://doi.org/10.1007/s11162-005-9003-2

Pyburn, D. T., Pazicni, S., Benassi, V. A., & Tappin, E. E. (2013). Assessing the relation between language comprehension and performance in general chemistry. *Chemistry Education Research and Practice*, *14*(4), 524–541. https://doi.org/10.1039/c3rp00014a

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, *13*(5783), 1–21.

Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS ONE*, *15*(9), 1–15. https://doi.org/10.1371/journal.pone.0238372

RealtimeBoard Inc. (2024). *Miro*. https://miro.com/about/

Roose, K. (2023). *Don't Ban ChatGPT in Schools. Teach With It.* The New York Times. https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). Sage Publications Inc.

Sarabipour, S., Niemi, N. M., Burgess, S. J., Smith, C. T., Filho, A. W. B., Ibrahim, A., & Clark, K. (2023). *Insights from a survey of mentorship experiences by graduate and postdoctoral researchers*. 1–38.

Schmid, K. M., & Wiles, J. R. (2022). *Call Her a Scientist: The Role of Mentors in Faculty Lab-Based Undergraduate Biology Research Experiences & Outcomes for Student Science Identity*. https://doi.org/10.1525/abt.2022.84.5.273

Shufflebarger, A. M., & Scott, K. A. (2020). Chemistry Departments Should Facilitate Graduate-Level Second-Language Writing Instruction. *Journal of Chemical Education*, *97*(12), 4220–4224. https://doi.org/10.1021/acs.jchemed.0c00896

Solano-Flores, G., & Trumbull, E. (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners. *Educational Researcher*, *32*(2), 3–13. https://doi.org/10.3102/0013189X032002003

Sotelo-Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. (2013). English language proficiency and test performance: an evaluation of bilingual students with the Woodcock-Johnson III Tests of Cogntitive Abilities. *Psychology in the Schools*, *50*(8), 781–797. https://doi.org/10.1002/pits

Stains, M. (2022). Keeping Up-to-Date with Chemical Education Research Standards. *Journal of Chemical Education*, *99*, 2213–2216. https://doi.org/10.1021/acs.jchemed.2c00488

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature*, *613*(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z

Subtirelu, N. C. (2015). She does have an accent but: Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society*, *44*(1), 35–62. https://doi.org/10.1017/S0047404514000736

Swanson, L. H., Bianchini, J. A., & Lee, J. S. (2014). Engaging in argument and communicating information: A case study of english language learners and their science teacher in an urban high school. *Journal of Research in Science Teaching*, *51*(1), 31–64. https://doi.org/10.1002/tea.21124

Talanquer, V. (2023). *Interview with the Chatbot: How Does It Reason?* https://doi.org/10.1021/acs.jchemed.3c00472

The National Academy of Sciences, National Academies of Science Engineering and Medicine, & The National Academy of Sciences. (2018). English Learners in STEM Subjects: Transforming Classrooms, Schools, and Lives. In D. Francis & A. Stephens (Eds.), *English*

Learners in STEM Subjects. The National Academies Press. https://doi.org/10.17226/25182

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Wiley, D. (2023). *What the Past Can Teach Us About the Future of AI and Education*. Campus Technology. https://campustechnology.com/articles/2023/04/06/what-the-past-can-teach-us-about-the-future-of-ai-and-education.aspx?s=ct_in_050523&oly_enc_id=0985D8346667B9T

Willis, G. (2011). Cognitive Interviewing in Practice: Think-Aloud, Verbal Probing, and Other Techniques. *Cognitive Interviewing*, 42–65. https://doi.org/10.4135/9781412983655.n4

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses. In *CRESST Report 732*.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly*, *39*(3), 347–352. https://doi.org/10.1598/RRQ.39.3.8

Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T., & Yaghi, O. M. (2023). ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *Journal of the American Chemical Society*, *145*(32), 18048–18062. https://doi.org/10.1021/jacs.3c05819

Chapter 8: Response process validity evidence in chemistry education research

8.1 Introduction to the research article presented in this chapter

Chapter 8 details a methods manuscript I co-authored with Nick Streja, a former Flynn group Master's student, and Dr. Flynn. The manuscript focuses on response process validity evidence, a type of validity evidence focused on individuals' interpretations of instrument prompts, items, *etc*. The manuscript has two parts. The first part, led by Nick, is a literature review of how response process validity evidence has been collected in different CER studies. The second part, led by myself, describes what response process validity evidence is, why and when it is important to collect in CER, and how to collect it.

Since its publication, this manuscript has emerged as a valuable resource for the CER community. The manuscript was featured in an editorial article by CER leader, Prof. Marilyne Staines, in her article "Keeping Up-to-Date with Chemical Education Research Standards", which outlines best practices from CER manuscripts submitted to the Journal of Chemical Education from 2020–2022 (Stains, 2022).

I want to thank and congratulate Nick, who started in the Flynn group around the same time I did. I always really enjoyed learning from and working with Nick, and our friendship made this project very fun and easy to work on.

8.2 References for Introduction

Stains, M. (2022). Keeping Up-to-Date with Chemical Education Research Standards. *Journal of Chemical Education*, *99*(6), 2213–2216. https://doi.org/10.1021/acs.jchemed.2c00488

8.3 Copyright

Reprinted from "Deng, J. M.; Streja, N.; Flynn, A. B. Response Process Validity Evidence in Chemistry Education Research. *J. Chem. Educ.* **2021**, *98* (12), 3656–3666. doi.org/10.1021/acs.jchemed.1c00749" with permission from the American Chemical Society and Division of Chemical Education, Inc. Copyright © 2021 American Chemical Society and Division of Chemical Education, Inc.

JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

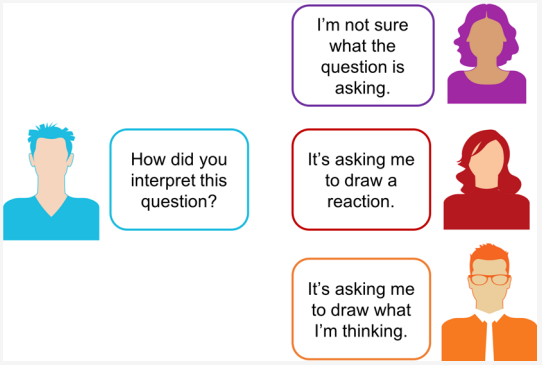
Response Process Validity Evidence in Chemistry Education Research

Jacky M. Deng, Nicholas Streja, and Alison B. Flynn*

Cite This: J. Chem. Educ. 2021, 98, 3656−3666 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Response process validity evidence can provide researchers with insight into how and why participants interpret items on instruments such as tests and questionnaires. In chemistry education research literature and the social sciences more broadly, response process validity evidence has been used and reported in a variety of ways. This paper's objective is to support researchers in developing purposeful, theory-driven protocols to investigate response processes. We highlight key considerations for researchers who are interested in using cognitive interviews in their research, including the following: the theoretical basis for response process investigations, collection of response process validity evidence through cognitive interviews, and use of that evidence to inform instrument modifications.



KEYWORDS: *General Public, Testing/Assessment, Chemical Education Research*

FEATURE: Chemical Education Research

Downloaded via UNIV OF OTTAWA on April 19, 2024 at 15:07:40 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Instruments are invaluable tools of measurement for researchers that allow us to directly and indirectly observe otherwise inaccessible phenomena.[1,2] For example, chemical instruments like spectrometers are regularly used to quantify physical and chemical entities and phenomena, with measurements providing the user with evidence related to the entities and phenomena being measured. Similarly, social science instruments (which include assessments, tests, surveys, and observation protocols) seek to measure variables within humans (respondents) directly and indirectly.[1,3] Among the goals of discipline-based education research (DBER), one is to "help identify and measure appropriate learning objectives and instructional approaches that advance students toward those objectives".[4] As DBER has grown, so too have the number of instruments designed to investigate different variables of interest.[5−10] Such instruments need to generate data of high validity and reliability. However, resources describing how to collect validity evidence in DBER contexts are limited.[11] This paper's objective is to support researchers in developing and reporting protocols to investigate a specific type of validity evidence: response process.

DEFINING VALIDITY

Validity and Types of Validity Evidence

Validity refers to the degree to which evidence/theory support interpretations of data for proposed uses with a specific population.[12−15] Collecting validity evidence is imperative in test development; without evidence for validity of the data obtained from a particular instrument in specific research context, there is uncertainty about what the instrument is actually measuring.[12]

In some cases, instruments may not yet exist for a particular goal, requiring instruments to be created and supported with appropriate validity evidence to determine whether they measure what they intend to measure. The effectiveness and suitability of social science instruments can vary depending on the goals and context of a particular evaluation. An instrument designed for a particular purpose in one context may need to be re-evaluated, reconfigured, or redesigned to provide meaningful measurements in pursuit of new research goals and/or within a new context. For example, an instrument designed to measure a particular variable within a university context may produce less valid data when used in a high school context.

Evidence for validity can be collected in several ways depending on the type, including construct validity, content validity, and criterion validity.[16] In this paper, we focus on construct validity only, which is an accumulation of evidence to support the interpretation of what a measure reflects.[17] Evidence for construct validity can be composed of evidence from

Received: July 7, 2021
Revised: November 5, 2021
Published: November 29, 2021



ACS Publications

© 2021 American Chemical Society and Division of Chemical Education, Inc.

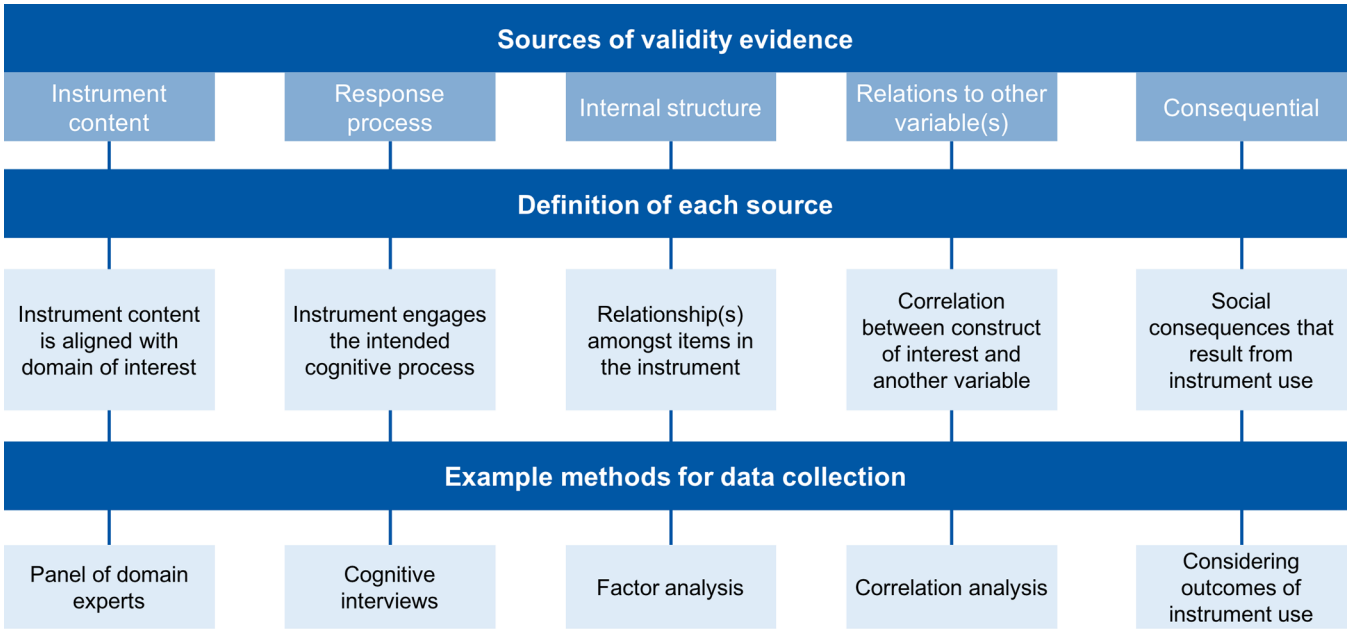


Figure 1. Sources of validity evidence, including definitions and examples of methods to collect each type of evidence.[12,13]

instrument content, response process, internal structure, relationships with other variables, and consequences of instrument use (Figure 1).[12,14] Each type of evidence can be collected using specific and established research methods; readers interested in detailed definitions of each of these types of validity evidence and their collection methods are referred to the *Standards for Educational and Psychological Testing*.[12]

Response Process Validity Evidence and Tourangeau's Cognitive Model

Response process represents one of the primary sources of validity evidence and is described as "evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinees".[12] By investigating respondents' response processes, researchers gain an understanding of how responses are generated, including what cognitive resources are activated, what barriers are faced by respondents, and whether interpretations of items differ from what developers intended.[18−20] In the context of instrument development, response process evidence can provide researchers with insight into whether the questions they are asking are (1) activating intended constructs and prompting respondents to respond with those constructs in mind and (2) structured in a fashion that allows respondents to appropriately interpret and respond.

Response process can be based in Tourangeau's four-stage cognitive model.[21,22] According to this model, in order to accurately respond to an item, a respondent must be able to (1) comprehend the item, (2) retrieve information relevant to the item's content, (3) make a judgment based upon the recall of knowledge relevant to the item, and (4) provide their answer within the given item structure. Breakdowns at each of these stages can be a result of various factors, including item wording, clarity, difficulty, and knowledge of relevant ideas held by respondents and/or their motivation.

The following is an example of how Tourangeau's model can be used to interpret response process data. A researcher interested in learning about students' self-efficacy in organic chemistry may design the item: "I am knowledgeable in acid—base chemistry", in which respondents are asked to respond on a Likert scale (e.g., strongly disagree to strongly agree). Tourangeau's model states that responding to this item requires respondents to engage in the following cognitive processes: (1) comprehending and interpreting critical terms and phrases within the item, such as what the terms "knowledgeable" and "acid−base chemistry" mean, (2) identifying and retrieving the necessary cognitive resources and information relevant to the item, (3) choosing to report an accurate estimation of their belief within the given item structure, and (4) selecting or producing an answer that is feasible within the given item structure. Evidence *for* response process validity would mean that the respondent's understanding of the item matches the item's intended meaning and that the respondent is able to report an accurate account of their response within the given item structure. Evidence *against* response process validity can manifest as difficulties that arise during at least one of Tourangeau's stages. For example, a respondent believing that "acid−base chemistry" is too ambiguous and broad a term to provide an accurate response would provide evidence *against* response process validity for this item (challenge experienced in stage 1 of Tourangeau's model). Other examples of evidence against response process validity could include difficulties with understanding the item as a whole (stage 1); the item activating irrelevant ideas within respondents (stage 2); respondents overestimating or underestimating their ability in their final response (stage 3); and/or respondents feeling as if they are not able to respond accurately within the test structure (stage 4).

As part of this work, we reviewed which stages of Tourangeau's model were probed in chemistry education research investigating response process; the details of this review will be discussed in the following section.

REVIEW OF RESPONSE PROCESS INVESTIGATIONS IN CHEMISTRY EDUCATION RESEARCH

To evaluate the state of how response process is investigated in chemistry education research, we conducted a literature search in *Journal of Chemical Education* (JCE) and *Chemistry Education*

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

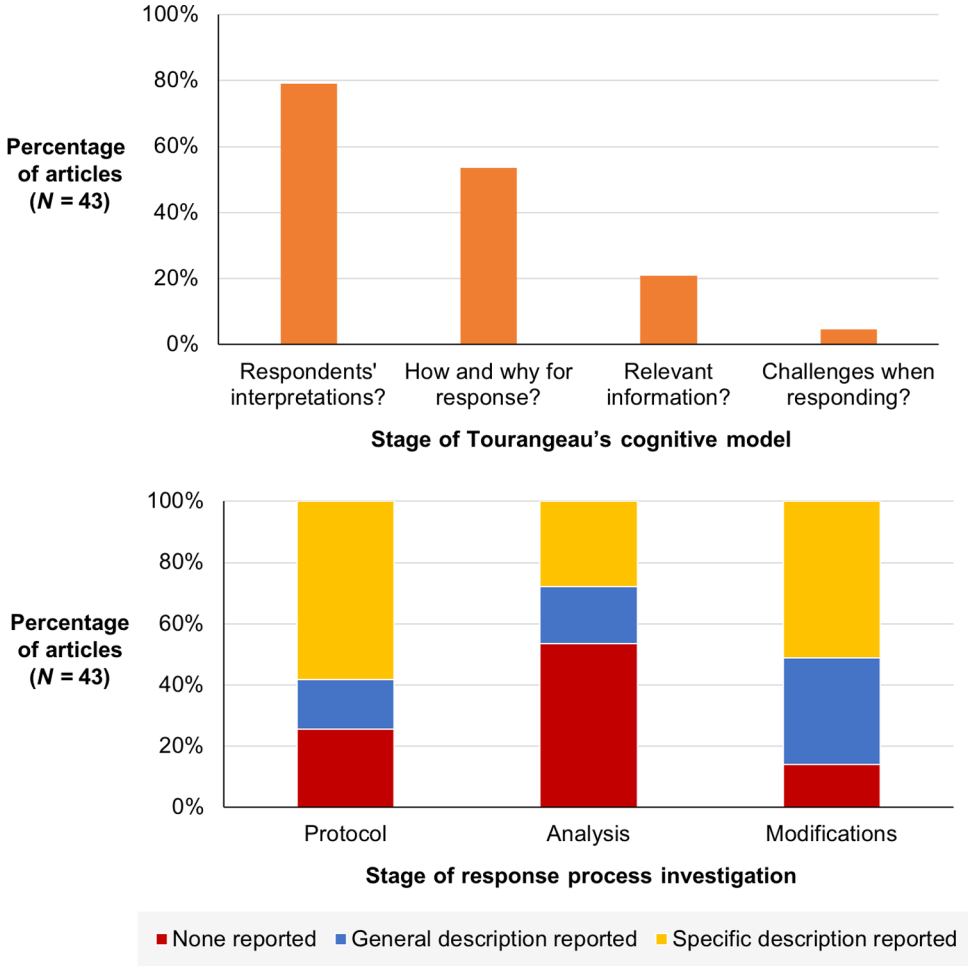


Figure 2. Summary of findings from a review of 43 articles describing collection of response process validity evidence in *Journal of Chemical Education* and *Chemistry Education Research and Practice*. Top: goals of response process investigations based on Tourangeau's cognitive model. Bottom: reporting of response process investigations). Ten articles reported that no modifications were needed.

Research and Practice (CERP) for any articles that used cognitive interviews or response process interviews to gather evidence for response process validity. We have provided the full details of the search, review, and analysis from the literature review in the Supporting Information. A total of 43 articles were included in our analysis, and we conducted a qualitative analysis of these 43 articles using *NVivo 12*. We searched the methods, results, and discussion sections from each article and identified where the authors described response process validity evidence. We then deductively coded these sections using the following four questions based on the four stages of Tourangeau's model: (1) How do you (the respondent) interpret the item? (2) What information do you believe to be relevant for responding to the item? (3) How and why did you decide on their given response? (4) What challenges did you face when providing their response within the given item structure, if any? We also reviewed articles for how response process validity evidence was collected, analyzed, and used to inform instrument modifications: none reported, general descriptions provided (i.e., the authors mention an analysis, but do not provide explicit details), and specific descriptions provided (i.e., the authors explain their process and resulting decisions with details). For modifications, we also captured whether the authors stated that no modifications were needed following response process investigations. Author N.S. conducted the analysis, and author J.M.D. reviewed author N.S.'s analysis for reliability. Any disagreements were discussed between the two authors prior to reaching consensus.

The findings from the 43 articles reviewed indicate that (1) a majority of CER investigations of response process focus on how respondents interpret and decide to respond to items and (2) 27 of the 43 articles did not report one (or more) of their response process protocol, analysis, and modifications, and 25 of the 43 reported at least one of these components generally (e.g., "semi-structured interviews were conducted" without specifying questions) (Figure 2). None of the articles explicitly referenced Tourangeau's model when describing their response process investigations or the rationale for investigations, which suggests that this model may be novel to the CER community. Our findings build on previous work; a 2013 study found that, of 20 instruments described in 37 publications, only four instruments reported response process validity evidence, with only two of these four providing specific details about protocols and the nature of changes to instruments due to response process evidence.[23] Our review of these same four articles noted similar results, with two of the studies only mentioning participants' interpretations and reporting general modifications.[24−27] The absence of response process investigations appears to be present in other research fields, as well. For example, a 2008 review of

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

Table 1. Potential Probes for Each Stage of Tourangeau's Four-Stage Model[22,41]

Stage of Tourangeau's Four-Stage Model	Potential Probes
Understanding: How does the respondent interpret the item?	Can you repeat the question in your own words?
	What does <term> mean to you in the context of the question?
	Was there anything confusing about the wording of the question?
	How would you rephrase the question to make it clearer?
Retrieval: What information does the respondent believe to be relevant for responding to the item?	How often have you thought about <content of question> before completing the question?
	How much do you know about <content of question>?
	Did you find the question easy/hard to answer? Why?
	What were you thinking about when you responded to the question?
Judgement: How and why does the respondent decide on their given response?	How comfortable are you with your response? Do you think it would change at all in a different setting?
	Did you feel pressured or pushed toward a specific answer? Why? Do you have any suggestions on we could change the question to correct for this?
	How did you decide to pick <response option selected>?
Response: What challenges did the respondent face then providing their response within the given item structure, if any?	How would your answer change, if at all, if this question was a <alternative response structure>? Do you prefer one or the other?
	For Likert scales: Here are a list of text labels. How would you rank them from lowest to highest?
	What would you change to the question to make responding easier or clearer?

validity investigations in psychology and education found response process was only described for 1.8% of measures.[28,29]

This work describes how researchers can design, enact, and report data using a common method for collecting response process evidence: cognitive interviews.[30−33] We include a discussion of different approaches to cognitive interviews, designing prompts to address different aspects of response process, and analyzing data from interviews to guide item/test modifications.

COGNITIVE INTERVIEWS TO COLLECT RESPONSE PROCESS VALIDITY EVIDENCE

Cognitive interviews can be used in instrument development to inform item revision decisions by providing evidence of validity based on both test content (alignment between test content and the intended domain and accessibility of test content to members of the intended population) and response process (thought processes involved in responding to an item).[12,22,34,35] When the goal is to collect evidence for response process validity, the focus of cognitive interviews (often called response process interviews) is to study how respondents think about and respond to materials presented to them, with a specific emphasis on what barriers and pitfalls arise during response process.[33,36] From these inquiries, qualitative data can be obtained that illuminate respondents' responses and behaviors, allowing researchers to determine whether item interpretations are (mis)aligned with developers' intentions and then to identify ways to modify items to improve validity.[37,38] When completed early in instrument development, cognitive interviews enable researchers to identify and correct issues in understanding and content before resources may be devoted to more extensive validity testing (e.g., factor analysis). Evidence of response process can range from collecting and indexing respondents' statements about their response to using theory to interpret respondents' statements in an effort to formulate generative mechanisms of how and why respondents decided to respond in the way(s) they did.

Designing Cognitive Interview Protocols

Interviews to gather response process evidence are a multistage process that can involve the following: identification of item intent, data collection, analysis, and comparison of respondent interpretation to intended meaning. Often, the main purpose of these interviews is to evaluate alignment between an item's intent and respondents' interpretations. Such interviews are distinct from interviews used to developing instrument items, such as using interview responses to develop concept inventory distractors.[9,39,40] In the first stage, researchers develop items for a particular instrument and document the intent of each item prior to cognitive interviewing, as well as any associated constructs the item is intended to investigate. These descriptions serve as the basis for comparison against respondents' interpretations.

Once items have been prepared, an interview protocol can be developed to gather information on respondents' response processes. A general framework based on Tourangeau's four-stage cognitive model of response process can be used to guide this process.[36] Table 1 gives examples of question to probe each stage of Tourangeau's cognitive model.

Of the 43 articles we reviewed, 34 focused on evaluating the readability and clarity of test items by asking students to read items aloud and then explain the meaning in their own words (Supporting Information). Other response process investigations have sought to develop rich, causal understandings of the how and why behind response processes through purposeful integration of theoretical and methodological frameworks. For example, Hosbein and Barbera (2020) asked respondents not only to read the item aloud and describe their interpretation, but also to choose an answer and describe why they chose the answer they did.[42] Of the 43 articles, 9 addressed the question "What information do you believe to be relevant to answering this question?" by probing during their analysis but do not give explicit information regarding how the probing was conducted.[25,43−50] Only two of the articles specifically addressed student challenges when answering questions, and only one explicitly asked students, "How difficult was this question to answer?"[46,51]

When designing cognitive interview protocols, researchers may also choose to focus not only cognitive processes, but also other, noncognitive processes involved in answering test items, such as self-referential, motivational, affective, situated, and dialogical response processes.[52] For example, methods can focus not only on what respondents are engaging with and why

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

Table 2. Summary of Qualitative Interview Approaches, Including Examples in a Response Process Interview Context[a]

Approach	Key Features	Advantages	Disadvantages
Informal conversational	No preplanned structure; researcher generates questions spontaneously within flow of conversation	Interview can be tailored to each participant and build off of previous interviews	Identifying trends in data can be difficult; increased risk of interviewer effects, so requires skilled interviewer
Interview guide	Specific questions/topics prepared before interview, with freedom to investigate spontaneous lines of inquiry	Researcher can focus conversation on several key topics, while still being flexible enough to maintain rapport and engage in details inquiry	Increased rigidity can limit the scope of the interview, such that unanticipated key perspectives could be missed
Standardized, open-ended	List of open-ended questions prepared before interview; researcher follows this list with limited deviation	Focused questioning ensures that all participants are asked the same questions, making data easier to analyze	Increasing rigidity further, so increased possibility that unanticipated key perspectives are missed
Fixed-response	List of fixed-response questions prepared before interview; researcher presents questions and options to interviewee	Fixing questioning and response options minimizes variability between responses; minimizes onus on interviewer to manage the conversation and often easier to collect data from larger samples to identify trends in response process	Very rigid; no opportunities to investigate response process perspectives beyond those provided to respondents

[a]Adapted from Patton (2015) and Willis (2011).

respondents are engaging in a particular response process cognitively, but also on how motivated respondents feel during the test-taking experience (e.g., fatigue) or have particular affective feelings toward particular items.

Approaches to Cognitive Interviews

Herein, we provide four general approaches to cognitive interviews (Table 2).[41,53] The four approaches can be broadly categorized within the continuum of think-aloud (TAP) and verbal probing protocols; they differ in terms of their relative flexibility/rigidity, resulting in distinct strengths and weaknesses depending on the goals of the research.[41]

The *informal conversational approach* is the most open-ended approach to interviewing. These types of interviews are characterized by their lack of preplanned structure and reliance on the researcher to generate questions within the flow of the interaction. The main strength of the conversational interview is its flexibility; questions can be specific to the interaction, allowing for deepened communication with the participant. Conversational interviews also allow the interviewer to take information from previous interviews to guide questioning in later interviews. The main weakness of informal conversational interviews is the difficulty involved in comparing and analyzing the data from interviews. The variability in questioning between interviews places the onus largely on the interviewer to generate useful data. The conversational interviewer must be skilled at generating rapid insights and questions. Because of their lack of structure, conversational approaches may be more susceptible to interviewer effects, leading questions, and biases, when unaccounted for, each threatening the integrity of the data collected from interviews.[54] Patterns and trends within the data set may also take longer to identify from conversational interviews, as responses from participants will vary as different participants are asked different sets of questions during their respective interviews.

Open-ended approaches often generate complex data from cognitive interviews, as researchers may be unsure where and how to make changes to an instrument if there is a large amount of variability in the topics discussed by interviewees. However, the flexibility associated with the informal conversational approach may be useful in research contexts where multiple rounds of interviews are feasible. For example, interviewers interested in where participants experience confusion when completing an instrument might conduct an initial round of interviews in an unstructured fashion to allow researchers to identify items that respondents express concerns about. Researchers can then use the information from these initial interviews to develop more structured, targeted questioning in future rounds of interviews.

The *interview guide approach* is characterized by researchers preparing a list of specific questions and/or issues to be explored prior to the interview. This approach ensures that each interviewee is confronted with the same general subjects of inquiry; however, the researcher maintains a degree of flexibility in this approach. One is free to formulate spontaneous responses to interviewee responses and establish a conversational approach within the confines of the predetermined subject of inquiry.

The main strength of the interview guide approach is that it provides the interviewer with a basic map or framework of how to facilitate the interview. This makes data collection more systematic across the sample. However, though there is some flexibility in this approach, the increased rigidity of the interview guide approach increases the risk of omitting important topics outside the purview of the protocol.

For response process investigations, the interview guide approach can be useful if the interviewer is interested in interviewees' thoughts and processes on specific instrument items. Researchers might have anticipated concerns about specific items, so they can focus their questioning during interviews on these items, while leaving room for interviews to discuss other issues and/or items. The interview guide approach can also be useful for cognitive interviews involving multiple rounds of interviews.[55] Researchers may choose to employ a conversational approach in initial rounds, allowing interviewees the freedom to discuss whatever concerns and interpretations come to mind during their response processes. Researchers could then identify any consistent concerns from these initial rounds, using this information to formulate an interview guide approach for subsequent rounds.

The *standardized open-ended approach* is characterized by careful and complete wording of each interview question prior to the interview. This approach imposes greater structure on the interview by not only fixing the general organization of the interview, but also fixing the wording of each question and the sequence of questions within the interview. However, questions are still open-ended, meaning that interviewees construct their own responses instead of choosing a response from a selection of predetermined possibilities. Like the interview guide approach, one of the strengths of the standardized open-ended approach is that interviewees answer the same questions, limiting the variability between interviews and allowing researchers to more easily organize/analyze the data and compare among

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

interviewees. However, the rigidity of this approach limits flexibility; the interviewer is granted few opportunities to pursue unanticipated topics or issues introduced during the interview, which can result in the loss of important data.

In cognitive interviews, a standardized open-ended approach may be useful if the interviewer wanted to learn about interviewees' perspectives on specific aspects of an instrument and/or a predetermined selection of instrument items. The researcher might identify a selection of items to discuss during the interview and ask each interviewee the same questions about each item. This would allow the researcher to gather targeted information about specific items and ensure that all interviewees have an opportunity to respond to the same set of questions during the interview.

In the *fixed-response approach*, questions and response options are determined in advance of the interview. Each interview follows exactly the same sequence of questions, and interviewees are instructed to select fixed responses instead of constructing their own. This approach allows for simpler data collection and analysis and also affords the interviewer with the opportunity to ask many questions in a short amount of time. However, the fixed-response approach constrains the opportunities afforded by qualitative inquiry. Because respondents must fit their thoughts and experiences within predetermined categories, the approach may misrepresent respondents' actual thoughts and experiences. Therefore, we suggest that researchers avoid relying on the fixed-response approach as the initial approach in gathering evidence for response process.

For cognitive interviews, the fixed-response approach may be useful in cases where researchers have already determined specific respondent concerns from initial rounds of interviews and want feedback on whether updates to items have remedied these specific concerns. For example, researchers can use an open approach in an initial round of interviews to determine common issues expressed by respondents and then make initial changes to test items to address these initial concerns. The researchers can then ask additional rounds of respondents whether these points of confusion persist or not using version of the instrument that multiple-choice options based on interviewees' concerns from the previous round (e.g., "Did you experience any of the following as you answered the item?").

The four approaches described above exist on a continuum, and researchers are certainly not limited to selecting only one of the approaches in their research. The approaches described can also be combined to develop a more comprehensive understanding of response process. For example, one study conducted open interviews with an initial sample of respondents to identify initial points for revision and then evaluated the impact of these revisions through targeted pilot administrations of their instrument.[56] Protocols can also be combined within a single interview; researchers might adopt a guided or standardized approach in the first part of the interview protocol to focus on specific items and then allow the interviewee to freely address any additional items or concerns during the latter part of the interview.

Concurrent and Retrospective Cognitive Interviews

Cognitive interviews can be conducted either concurrently or retrospectively. The two approaches differ in recall accuracy and disruption of response process. Concurrent interviews require interviewees to verbalize their experiences and response processes as they work through test items. For example, in one study, researchers provided participants with a copy of their test and, as they worked through the test, asked each participant to read each item aloud, explain what was being asked, choose an answer, and explain their choice.[9] Because interviewees were being asked about their cognitive processes *while* they went through the test, interviewees' descriptions of their processes should have been more accurate than if the interviewee had been recalling their mental processes at a later point in time. However, this approach comes at the risk of disrupting participants' thinking during the test-taking experience, meaning that the response process expressed by the interviewee may be less similar to the response process of an actual test respondent.[41] Asking interviewees to verbalize their response process may increase cognitive demand and introduce response process disruptions that would not be encountered if respondents completed the test on their own.

In contrast, retrospective interviews are conducted after the interviewee has had a chance to complete the instrument without disruption. For example, in another study, researchers asked respondents to complete the survey in full before engaging in a questioning period.[26] Compared to concurrent protocols, retrospective interviews ease the cognitive demand on interviewees during test-taking, allowing them to complete the test without disruption. However, interviewees may be more likely to forget and introduce bias as they try to recall their response process at a later point in time.

One way to mitigate the challenges associated with each of these approaches is to combine the two approaches. One study conducted interviews that incorporated both concurrent and retrospective components;[56] in this case, respondents were asked to actively think-aloud as they answered test items (concurrent). After they had completed the items, the researchers then asked respondents about their interpretations of test items (retrospective). Another alternative to reduce recall bias is retrospective interviewing at the item level, as opposed to the whole test/instrument level. This approach can be especially helpful for open-ended or multiple-choice response options (e.g., concept inventory items) as opposed to affective statements (e.g., self-efficacy beliefs) on a Likert scale.

Sampling

Validity is a body of evidence that supports the use of data for a particular purpose with a specific population. This body of evidence is collected by an instrument for a specific target population, so participants invited to cognitive interviews should be as closely aligned as possible with members of the intended test-taking population.[12] Cognitive interviews should seek a diversity of perspectives; therefore, we suggest that sampling for cognitive interviews is most effective when restricted to the test-taking population, but recruitment from this population is random and/or purposeful with diversity in mind. As with any research with human subjects, Research Ethics Board approval or exemption must be obtained before recruiting for and conducting cognitive interviews.

Cognitive interviews are a qualitative method of inquiry, so sample sizes will vary depending on the specific goals of investigations. From CER literature describing the use of cognitive interviews, sample sizes have ranged from 2 to 58 (Supporting Information). Recommendations from other sources range from 5 to 15,[22,57,58] with some research suggesting that increasing the sample sizes beyond this range does not seem to uncover new response process problems.[59]

Using sequential rounds of interviewing can be advantageous when collecting evidence of response process,[33,55] as some have

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

Table 3. Example of the Cognitive Interview Process to Probe How/Why Respondents Decide on Their Response[a]

Stage 1: Prepare Item Intent and Any Associated Construct Dimensions	Stage 2: Develop Interview Protocol	Stage 3: Conduct Cognitive Interviews	Stage 4: Analysis of Interview Data	Stage 5: Compare to Intent and Dimensions; Decide on Modification Strategy
Item: Rate your confidence to propose a possible mechanism for a reaction, given the starting materials and products (1, not at all confident, to 7, completely confident)	**Stage of Tourangeau's model to be probed:** Judgement	**Respondent 1**: "I am confident I can propose a mechanism for a reaction from this course (organic chemistry II) but I don't know if it means for *any* mechanism or just the ones that were taught."	**Key phrases**: "from this course"; "I don't know if it means *any*"; "one that I've seen before"	**Conclusion**: The item's intention was to assess respondents' self-efficacy beliefs related to organic mechanisms. Data show that the item is too vague for respondents to clearly understand the context of what is being asked.
Construct: self-efficacy	**Probe**: How did you decide on your response?	**Respondent 2**: "It depends on the mechanism. If it's one I'm not familiar with then I wouldn't be very confident but if it's one that I've seen I would be more confident."	**Summary**: Respondents are unsure about the context of the question, which could indicate the item is too vague.	**Recommendation**: Revise the item to make the context more specific. For example, "Propose a possible mechanism for the reactions taught in this course, given the starting materials and products." Conduct another round of cognitive interviews to investigate the impact of modification.
Intent: Assess respondent's perception of self-efficacy beliefs related to organic reaction mechanisms taught in an organic chemistry course	Why did you select the response that you did?			
	Do you think your response would change at all in a different setting?			

[a]Adapted from Peterson et al., 2017.

done in CER.[9,40,56] This strategy is particularly useful in instrument development because it provides researchers an opportunity to investigate the impact of instrument modifications following initial interviews. For example, researchers may interview five respondents and make changes to an instrument based on the evidence gathered from this first round. To better understand the impact of these changes, the researcher can then conduct a second round of interviews with a new set of respondents to identify (1) whether instrument modifications have resolved or exacerbated issues from the first round and (2) if there are any new challenges faced by respondents in the second round of interviews. For example, Stephenson et al. (2020) conducted a round of interviews to gather initial data about respondents' interpretations of items (n = 5; round 1), revised items based on these interviews, conducted a second round of response process investigation by administering the revised instrument to members of the target population (n = 30; round 2), and used the sum of the findings to develop their final task version.[56] However, conducting interviews in rounds can mean that the final sample size for a response process investigation may not be known at the beginning of the study. Researchers may begin their research with a specific sample size in mind, but that number can be adjusted on the basis of what is learned from interviews, as inquiry deepens, or when data saturation is observed.[60]

ANALYZING COGNITIVE INTERVIEW DATA TO INFORM INSTRUMENT MODIFICATIONS

Of the three stages of response process investigations, analysis of response process data was found to have the highest frequency of "*none reported*". Variable reporting of response process data analysis has also been reported in other reviews of response process investigations in social science literature.[28,29,61] There is variation in analysis methods for cognitive interview data.[22] Analysis of cognitive interview data can be less intensive, characterized by interviewers taking notes of problems during the interview process and comparing notes across interviews.[62] For example, one analysis is described as follows: "If the rationale students provided was not aligned with the researcher's intended meaning, the item was flagged for modification or removal. If students gave a rationale that was aligned with the intended meaning, there was validity evidence that the item was being interpreted as intended and could be used in future implementations of the scale in the given context."[63] In contrast, analyses involving coding schemes and classifying response process issues taxonomically have also been employed.[55,64] For example, one analysis stated that "audio files of the interviews were transcribed by a commercial transcription service. All analysis was conducted on the interview transcripts, led by author Hosbein. Cognitive interview analysis was used by Hosbein to establish a coding rubric. This entailed carefully reading each transcript and documenting the language students used in explaining their responses. These reviews were used to establish initial codes regarding students' scale-based language... Hosbein and a secondary (nonauthor) researcher independently coded all 40 transcripts using the initial rubric. After discussing discrepancies in coding, the rubric was revised and additional transcripts were independently recoded. The final rubric had an acceptable Cohen's kappa value of 0.8 and was used to recode the remaining transcripts."[48] Research on the range of analytical methods employed has identified benefits and challenges with each approach. Less intensive analyses unearth fewer response problems and may be more prone to bias,[62,65] while more intensive analyses may not be feasible. One may simply lack the resources (e.g., time, funding) to engage in a full qualitative analysis of response process data, especially if instrument administration is time-dependent.[55] In response to the ambiguity in the literature about what is "best" and to serve a range of research contexts, researchers have devised qualitative methods that seek to balance feasibility and reliability;[22,34] in Table 3, we provide an example for probing item comprehension to support researchers in a variety of contexts.

Regardless of context, we recommend that authors of future studies *justify* the depth and breadth of the methods they use based on the *intentions* and/or *limitations* of the response process investigation.[66] Though there is no single or "best" approach for analyzing and modifying items based on cognitive interview data, readers will benefit from the explicit descriptions of such investigations whenever possible. To best highlight how and

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

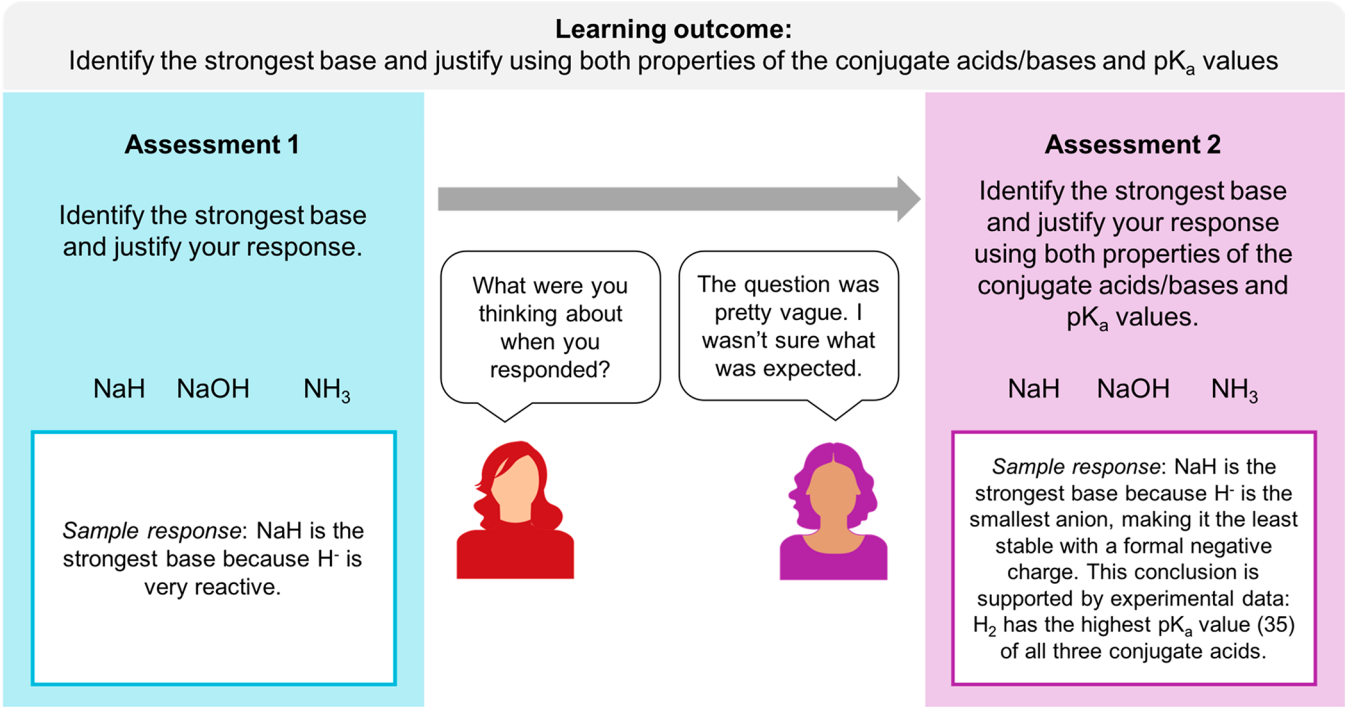


Figure 3. Example of revising assessment items based on response process evidence gathered from students' feedback.

why students interpret items during instrument development/modification for the research community, reports should explicitly describe:

(1) Who is sampled to participate in any response process investigations?
(2) What protocols are enacted and what is the rationale for these protocols?
(3) Why are data analyzed and used to inform instrument modifications (if at all)?

■ ALTERNATIVES TO COGNITIVE INTERVIEWS

Interviews in particular can be resource-intensive and subject to threats and limitations such as interviewer effects, disproportionately motivated participants, and small sample sizes. Although cognitive interviews have been the predominant method to investigate response process throughout the social sciences and discipline-based education research, other methods exist that may be more appropriate and feasible for a particular context. Alternatives that can be used instead of or in parallel with cognitive interviewing to develop more comprehensive understandings of response processes include open response questions, vignettes, focus groups, and eye-tracking. Details about these alternatives can be found in the Supporting Information.

■ COLLECTING RESPONSE PROCESS VALIDITY EVIDENCE IN TEACHING

Course assessments are instruments that are designed to elicit evidence of students' knowledge, skills, or values based on a specific set of learning outcomes within a specific learning context. Unlike instruments used in research, course assessments may also be used for formative purposes, providing educators and learners with "data" that can be used to better support student learning within a course, guide learning strategies, etc. Like all instruments, one can collect evidence for validity for course assessments, helping ensure that inferences about student achievement of learning outcomes are as accurate as possible.[12,67] However, the rigor of validity investigations (in both research and teaching) will depend greatly on feasibility, goals, context, etc. Accordingly, the following section describes methods for how response process could be investigated on course assessments while balancing rigor and feasibility.

Educators may speak with several students to gain insight into students' response processes, similar to the cognitive interviews described above. Educators may also provide opportunities for students to describe how they arrived at their answer or how they approached specific questions on an assessment (e.g., providing an option to write out their thinking after a multiple-choice question).[67−69] By analyzing/evaluating students' responses, educators gain insight into students' response processes, which can inform how to evaluate responses, tailor instruction, and improve assessments. Educators can also ask similar types of questions throughout the sequence of a course and refine the questions over time. For example, educators can ask a question across different formative and summative assessments and review student responses to refine the question as the course progresses, similar to conducting cognitive interviews in rounds (Figure 3).[70,71]

■ CONCLUSIONS

In this paper, we highlight key considerations for using cognitive interviews to enhance validity, including: (1) how Tourangeau's four-stage cognitive model can be used to guide response process investigations,[36] (2) different approaches to cognitive interviewing and the advantages/challenges associated with each, and (3) suggested practices when analyzing and using cognitive interview data to inform instrument modifications. Cognitive interviews can contribute to the piece of an instrument's overall validity argument by providing evidence of respondents' response processes. We examined 43 CER

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

articles in which the authors used cognitive interviews to collect response process validity evidence. The majority of articles described how/why respondents decided on their responses to items and respondents' interpretations of items. Response process investigation reporting varied among articles. Although the variability is unsurprising given the varied goals and contexts for each investigation, 27 of the 43 articles did not report one (or more) of their response process protocols, analyses, and modifications. Transparent and detailed reporting of response process validity methods, analyses, results, and modifications will help readers understand the context of the investigations and determine the relevance of the works, although there will understandably be variation in the descriptions based on context and purpose. We hope that this paper supports researchers in improving the validity of data obtained from research and educational instruments, developing and reporting purposeful protocols to collect and refine instruments based on evidence on students' response processes, and ultimately improving the quality of data collection toward improving research outcomes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00749.

> Summary of response process validity reporting in CER articles, a list of alternative methods for collecting response process validity evidence, description of literature review method and analysis, additional figures from the literature review, and details on conducting interviews remotely (PDF, DOCX)

■ AUTHOR INFORMATION

Corresponding Author

Alison B. Flynn − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada;* orcid.org/0000-0002-9240-1287; Email: alison.flynn@uottawa.ca

Authors

Jacky M. Deng − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada;* orcid.org/0000-0002-2633-6624

Nicholas Streja − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.1c00749

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Author J.M.D. would like to thank the Social Sciences and Humanities Research Council (SSHRC) for funding through the Vanier Canada Graduate Scholarship.

■ REFERENCES

(1) Opfer, J. E.; Nehm, R. H.; Ha, M. Cognitive Foundations for Science Assessment Design: Knowing What Students Know about Evolution. *J. Res. Sci. Teach.* **2012**, *49* (6), 744−777.

(2) Campbell, C. E.; Nehm, R. H. A Critical Analysis of Assessment Quality in Genomics and Bioinformatics Education Research. *CBE Life Sci. Educ.* **2013**, *12* (3), 530−541.

(3) Bollen, K. A. Latent Variables in Psychology and the Social Sciences. *Annu. Rev. Psychol.* **2002**, *53* (1), 605−634.

(4) National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; Washington, DC, 2012.

(5) Sirum, K.; Humburg, J. The Experimental Design Ability Test (EDAT). *Bioscene* **2011**, *37* (1), 9−16.

(6) Smith, M. K.; Wood, W. B.; Knight, J. K. The Genetics Concept Assessment: A New Concept Inventory for Gauging Student Understanding of Genetics. *CBE Life Sci. Educ.* **2008**, *7* (4), 422−430.

(7) Denofrio, L. A.; Russell, B.; Lopatto, D.; Lu, Y. Mentoring: Linking Student Interests to Science Curricula. *Science (Washington, DC, U. S.)* **2007**, *318* (5858), 1872−1873.

(8) Smith, M. K.; Jones, F. H. M.; Gilbert, S. L.; Wieman, C. E. The Classroom Observation Protocol for Undergraduate Stem (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE Life Sci. Educ.* **2013**, *12* (4), 618−627.

(9) Atkinson, M. B.; Popova, M.; Croisant, M.; Reed, D. J.; Bretz, S. L. Development of the Reaction Coordinate Diagram Inventory: Measuring Student Thinking and Confidence. *J. Chem. Educ.* **2020**, *97* (7), 1841−1851.

(10) Wren, D.; Barbera, J. Gathering Evidence for Validity during the Design, Development, and Qualitative Evaluation of Thermochemistry Concept Inventory Items. *J. Chem. Educ.* **2013**, *90*, 1590.

(11) Reeves, T. D.; Marbach-Ad, G. Contemporary Test Validity in Theory and Practice: A Primer for Discipline-Based Education Researchers. *CBE Life Sci. Educ.* **2016**, *15* (1), rm1.

(12) American Psychological Association; American Educational Research Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, 2014.

(13) Kane, M. T. Validating the Interpretations and Uses of Test Scores. *J. Educ. Meas.* **2013**, *50* (1), 1−73.

(14) Messick, S. Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performance as Scientific Inquiry into Score Meaning. *Am. Psychol.* **1995**, *50*, 741−749.

(15) Cronbach, L. J. Test Validation. In *Educational Measurement*; Thorndike, R. L., Ed.; American Council on Education: Washington, DC, 1971; pp 443−507.

(16) Guion, R. M. On Trinitarian Doctrines of Validity. *Prof. Psychol.* **1980**, *11* (3), 385−398.

(17) Cronbach, L. J.; Meehl, P. E. Construct Validity in Psychological Tests. *Psychol. Bull.* **1955**, *52*, 281−302.

(18) Hughes, K. Comparing Pretesting Methods: Cognitive Interviews, Respondent Debriefing, and Behavior Coding. *Surv. Methodol.* **2004**, *2*, 1−20.

(19) Presser, S.; Couper, M. P.; Lessler, J. T.; Martin, E.; Martin, J.; Rothgeb, J. M.; Singer, E. Methods for Testing and Evaluating Survey Questions. *Public Opin. Q.* **2004**, *68* (1), 109−130.

(20) Ryan, K.; Gannon-Slater, N.; Culbertson, M. J. Improving Survey Methods With Cognitive Interviews in Small- and Medium-Scale Evaluations. *Am. J. Eval.* **2012**, *33* (3), 414−430.

(21) Tourangeau, R. Cognitive Sciences and Survey Methods. In *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*; National Academies Press: Washington, DC, 1984; pp 73−100.

(22) Peterson, C. H.; Peterson, N. A.; Powell, K. G. Cognitive Interviewing for Item Development: Validity Evidence Based on Content and Response Processes. *Meas. Eval. Couns. Dev.* **2017**, *50* (4), 217−223.

(23) Arjoon, J. A.; Xu, X.; Lewis, J. E. Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *J. Chem. Educ.* **2013**, *90* (5), 536−545.

(24) Mulford, D. R.; Robinson, W. R. An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *J. Chem. Educ.* **2002**, *79* (6), 739.

(25) Oloruntegbe, K. O.; Ikpe, A. Ecocultural Factors in Students' Ability to Relate Science Concepts Learned at School and Experienced at Home: Implications for Chemistry Education. *J. Chem. Educ.* **2011**, *88* (3), 266−271.
(26) Barbera, J.; Adams, W. K.; Wieman, C. E.; Perkins, K. K. Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *J. Chem. Educ.* **2008**, *85* (10), 1435−1439.
(27) Stains, M.; Escriu-Sune, M.; Molina Alvarez de Santizo, M. L.; Sevian, H. Assessing Secondary and College Students' Implicit Assumptions about the Particulate Nature of Matter: Development and Validation of the Structure and Motion of Matter Survey. *J. Chem. Educ.* **2011**, *88* (10), 1359−1365.
(28) Cizek, G. J.; Bowen, D.; Church, K. Sources of Validity Evidence for Educational and Psychological Tests: A Follow-up Study. *Educ. Psychol. Meas.* **2010**, *70* (5), 732−743.
(29) Cizek, G. J.; Rosenberg, S. L.; Koons, H. H. Sources of Validity Evidence for Educational and Psychological Tests. *Educ. Psychol. Meas.* **2008**, *68* (3), 397−412.
(30) DeWalt, D. A.; Rothrock, N.; Yount, S.; Stone, A. A. Evaluation of Item Candidates: The PROMIS Qualitative Item Review. *Med. Care* **2007**, *45* (5), S12−S21.
(31) Irwin, D. E.; Varni, J. W.; Yeatts, K.; DeWalt, D. A. Cognitive Interviewing Methodology in the Development of a Pediatric Item Bank: A Patient Reported Outcomes Measurement Information System (PROMIS) Study. *Health Qual. Life Outcomes* **2009**, *7*, 3.
(32) Poole, H. M.; Murphy, P.; Nurmikko, T. J. Development and Preliminary Validation of the NePIQoL: A Quality-of-Life Measure for Neuropathic Pain. *J. Pain Symptom Manage.* **2009**, *37* (2), 233−245.
(33) Willis, G. Introduction to Cognitive Interviewing. *Cogn. Interviewing* **2005**, 2−11.
(34) Castillo-Díaz, M.; Padilla, J. L. How Cognitive Interviewing Can Provide Validity Evidence of the Response Processes to Scale Items. *Soc. Indic. Res.* **2013**, *114* (3), 963−975.
(35) Haladyna, T. M. *Developing and Validating Test Items*; Routledge: New York, 2013. DOI: 10.4324/9780203850381
(36) Tourangeau, R.; Rips, L. J.; Rasinski, K. *The Psychology of Survey Response*; Cambridge University Press: New York, 2000.
(37) Fowler, F. J. *Survey Research Methods*; Sage Publications Inc.: Thousand Oaks, CA, 2013.
(38) Biemer, P. P.; Lyberg, L. E. *Introduction to Survey Quality*; John Wiley & Sons, Inc.: Hoboken, NJ, 2003.
(39) Roche Allred, Z. D.; Bretz, S. L. Development of the Quantization and Probability Representations Inventory as a Measure of Students' Understandings of Particulate and Symbolic Representations of Electron Structure. *J. Chem. Educ.* **2019**, *96* (8), 1558−1570.
(40) Abell, T. N.; Bretz, S. L. Development of the Enthalpy and Entropy in Dissolution and Precipitation Inventory. *J. Chem. Educ.* **2019**, *96*, 1804.
(41) Willis, G. Cognitive Interviewing in Practice: Think-Aloud, Verbal Probing, and Other Techniques. *Cogn. Interviewing* **2005**, 42−65.
(42) Hosbein, K. N.; Barbera, J. Development and Evaluation of Novel Science and Chemistry Identity Measures. *Chem. Educ. Res. Pract.* **2020**, *21* (3), 852−877.
(43) Galloway, K. R.; Bretz, S. L. Development of an Assessment Tool to Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2015**, *92* (7), 1149−1158.
(44) Versprille, A.; Zabih, A.; Holme, T. A.; McKenzie, L.; Mahaffy, P.; Martin, B.; Towns, M. Assessing Student Knowledge of Chemistry and Climate Science Concepts Associated with Climate Change: Resources To Inform Teaching and Learning. *J. Chem. Educ.* **2017**, *94* (4), 407−417.
(45) Hale, L. V. A.; Lutter, J. C.; Shultz, G. V. The Development of a Tool for Measuring Graduate Students' Topic Specific Pedagogical Content Knowledge of Thin Layer Chromatography. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 700−710.
(46) Harsh, J. A. Designing Performance-Based Measures to Assess the Scientific Thinking Skills of Chemistry Undergraduate Researchers. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 808−817.
(47) Lutter, J. C.; Hale, L. V. A.; Shultz, G. V. Unpacking Graduate Students' Knowledge for Teaching Solution Chemistry Concepts. *Chem. Educ. Res. Pract.* **2019**, *20* (1), 258−269.
(48) Komperda, R.; Hosbein, K. N.; Phillips, M. M.; Barbera, J. Investigation of Evidence for the Internal Structure of a Modified Science Motivation Questionnaire II (MSMQ II): A Failed Attempt to Improve Instrument Functioning across Course, Subject, and Wording Variants. *Chem. Educ. Res. Pract.* **2020**, *21* (3), 893−907.
(49) Lazenby, K.; Becker, N. M. Evaluation of the Students' Understanding of Models in Science (SUMS) for Use in Undergraduate Chemistry. *Chem. Educ. Res. Pract.* **2021**, *22* (1), 62−76.
(50) Brandriet, A. R.; Bretz, S. L. The Development of the Redox Concept Inventory as a Measure of Students' Symbolic and Particulate Redox Understandings and Confidence. *J. Chem. Educ.* **2014**, *91* (8), 1132−1144.
(51) Villalta-Cerdas, A.; Sandi-Urena, S. Self-Explaining Effect in General Chemistry Instruction: Eliciting Overt Categorical Behaviours by Design. *Chem. Educ. Res. Pract.* **2014**, *15* (4), 530−540.
(52) Sorana Laneanu, M.; Hubley, A. M. A Model Building Approach to Examining Response Processes as a Source of Validity Evidence for Self-Report Items and Measures. In *Understanding and Investigating Response Processes in Validation Research*; Zumbo, B. D.; Hubley, A. M., Eds.; Springer International Publishing: Vancouver, BC, 2017; pp 115−136.
(53) Patton, M. Q. Distinguishing Interview Approaches and Types of Interviews. In *Qualitative Research & Evaluation Methods*; Sage, 2015; pp 641−648.
(54) Kreuter, F. Interviewer Effects. In *Encyclopedia of Survey Research Methods*; Lavrakas, P. J., Ed.; Sage Publications Inc.: Thousand Oaks, CA, 2011; pp 370−371.
(55) Conrad, F. G.; Blair, J. Data Quality in Cognitive Interviews: The Case of Verbal Reports. In *Methods for Testing and Evaluating Survey Questionnaires*; Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Singer, E., Eds.; John Wiley & Sons, Inc., 2004; pp 67−87.
(56) Stephenson, N. S.; Duffy, E. M.; Day, E. L.; Padilla, K.; Herrington, D. G.; Cooper, M. M.; Carmel, J. H. Development and Validation of Scientific Practices Assessment Tasks for the General Chemistry Laboratory. *J. Chem. Educ.* **2020**, *97* (4), 884−893.
(57) Beatty, P. C.; Willis, G. B. Research Synthesis: The Practice of Cognitive Interviewing. *Public Opin. Q.* **2007**, *71* (2), 287−311.
(58) Cresswell, J. W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*; Pearson: Boston, MA, 2012.
(59) Blair, J.; Conrad, F. G. Sample Size for Cognitive Interview Pretesting. *Public Opin. Q.* **2011**, *75* (4), 636−658.
(60) Patton, M. Q. Sample Size for Qualitative Designs. In *Qualitative Research & Evaluation Methods*; Sage Publications Inc., 2015; pp 470−477.
(61) Zumbo, B. D.; Shear, B. R. *The Concept of Validity and Some Novel Validation Methods*; Northeastern Educational Research Association: Rocky Hill, CT, 2011; p 56.
(62) Willis, G. Analyzing and Documenting Cognitive Interview Results. *Cognitive Interviewing*; Sage, 2011; pp 151−175. DOI: 10.4135/9781412983655.n11
(63) Hensen, C.; Barbera, J. Assessing Affective Differences between a Virtual General Chemistry Experiment and a Similar Hands-On Experiment. *J. Chem. Educ.* **2019**, *96* (10), 2097−2108.
(64) Woolley, M. E.; Bowen, G. L.; Bowen, N. K. The Development and Evaluation of Procedures to Assess Child Self-Report Item Validity. *Educ. Psychol. Meas.* **2006**, *66*, 687−700.
(65) DeMaio, T. J.; Landreth, A. Do Different Cognitive Interview Techniques Produce Different Results? *Methods Test. Eval. Surv. Quest.* **2004**, 89−108.

(66) Smith, K. C.; Alonso, V. Measuring Student Engagement in the Undergraduate General Chemistry Laboratory. *Chem. Educ. Res. Pract.* **2020**, *21* (1), 399−411.
(67) Bonner, S. M. Validity in Classroom Assessments: Purpose, Properties, and Principles. In *SAGE Handbook of Research on Classroom Assessment*; McMillan, J. H., Ed.; Sage Publications Inc.: Thousand Oaks, CA, 2013; pp 87−106.
(68) Yale Poorvu Centre for Teaching and Learning. Considering Validity in Assessment Design. https://ctl.yale.edu/sites/default/files/basic-page-supplementary-materials-files/considering_validity_in_assessment_design_handout.pdf (accessed Jul 5, 2021).
(69) Talbot, R. M. On the Importance of the Validity of Classroom Assessments. https://medium.com/@Bud_T/on-the-importance-of-the-validity-of-classroom-assessments-fe019f3bdd8c (accessed Jul 5, 2021).
(70) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821−1832.
(71) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chem. Teach. Int.* **2021**, *3* (1), 19−30.

https://doi.org/10.1021/acs.jchemed.1c00749
J. Chem. Educ. 2021, 98, 3656−3666

Chapter 9: Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities

9.1 Introduction to the research article presented in this chapter

The following chapters (Chapters 9–12) describe projects I contributed to or led that were *independent* from my research with the Flynn group. I justify their inclusion in Section 1.3 of Chapter 1.

Chapter 9 describes a manuscript I led in collaboration with graduate students across North America: Leah McMunn (University of Ottawa), Meagan Oakley (University of Alberta), Hoang Dang (University of Iowa), and Becky Rodriguez (University of Minnesota). The idea for the manuscript emerged during the 2020 global anti-racism movement—we came together with a mutual desire to contribute to ongoing conversations and initiatives to improve equity and inclusion in chemistry and science. The manuscript details the various equity, diversity, and inclusion (EDI) initiatives we had each led or contributed as graduate students at our respective institutions, with their benefits and impacts justified using the community cultural wealth (CCW) framework. The explicit goal of the work was to provide readers an overview of different initiatives they might implement at their own institutions, including their potential benefits and limitations. With reflection and time, I think also we had implicit goals, to: (1) demonstrate to readers that graduate students can lead (and have led) conversations and initiatives about EDI, and (2) to inspire other graduate students to effect change at their own institutions.

I'm proud of this project for several reasons. First, it was my first project/manuscript as corresponding author. I learned a lot about leadership during this process, including what I was good at and what I needed to work on. Second, this work was published in a Special Issue of the *Journal of Chemical Education* focused on Diversity, Equity, Inclusion, and Respect. Our article was one of just a few published by students for this special issue.

Thank you to Leah, Meagan, Hoang, and Becky. I look up to and admire each of you, so I'm humbled that you believed in me to lead this project.

9.2 Copyright

Reprinted from "Deng, J. M.; McMunn, L. E.; Oakley, M. S.; Dang, H. T.; Rodriguez, R. S. Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities. *J. Chem. Educ.* **2022**, *99* (1), 373–382. doi.org/10.1021/acs.jchemed.1c00485" with permission from the American Chemical Society and Division of Chemical Education, Inc. Copyright © 2022 American Chemical Society and Division of Chemical Education, Inc.

JOURNAL OF CHEMICAL EDUCATION

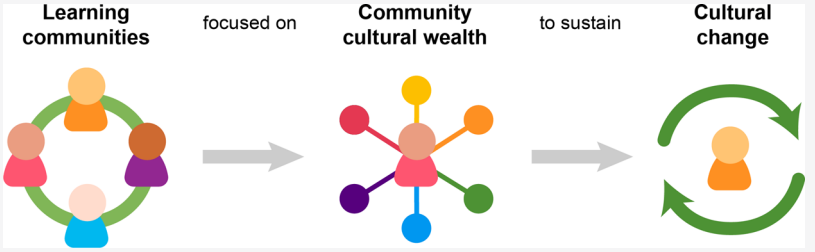
pubs.acs.org/jchemeduc Article

Toward Sustained Cultural Change through Chemistry Graduate Student Diversity, Equity, and Inclusion Communities

Jacky M. Deng,* Leah E. McMunn,⊥ Meagan S. Oakley,⊥ Hoang T. Dang,⊥ and Rebeca S. Rodriguez⊥

Cite This: J. Chem. Educ. 2022, 99, 373−382 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Evidence from the past several decades has repeatedly found that the chemical enterprise presents systemic barriers for people with marginalized identities. Initiatives and actions that remove systemic barriers and support the success of all students, especially those from equity-deserving groups, are essential. However, fostering a culture of inclusion requires actions that focus on not only immediate outcomes but also sustained and lasting impacts. Chemistry graduate student communities focused on diversity, equity, inclusion, and respect (DEIR) can contribute to the transformation of our shared chemistry community into one where DEIR principles are foundational. Chemistry graduate student DEIR communities (CGDEIRCs) are learning communities in which students from equity-deserving groups in chemistry gain various forms of cultural wealth that support their attainment of cultural capital. By supporting students' development of skills, networks, and resources to attain success, over time, CGDEIRCs may contribute to both short- and long-term DEIR cultural shifts in the broader chemistry community.

KEYWORDS: *Graduate Education/Research, Minorities in Chemistry, Professional Development, Women in Chemistry, Collaborative/Cooperative Learning, Public Understanding/Outreach, Interdisciplinary/Multidisciplinary*

Downloaded via UNIV OF OTTAWA on April 18, 2024 at 21:45:42 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

In the summer of 2020, heightened awareness of acts of police brutality against Black Americans sparked protests around the world calling for racial justice. The impact of these protests was widespread and resulted in discussions centered on diversity, equity, inclusion, and respect (DEIR) in various sectors. In science, technology, engineering, and mathematics (STEM) and STEM education, many of these conversations prompted our communities to reflect on the various systemic barriers experienced by marginalized groups and ways to better support individuals from these groups in STEM programs and careers. Over the past few months, various institutions have released statements and plans calling for people to reflect, learn, and act in ways that would improve the state of DEIR in STEM. Action is certainly needed in the present, but actions must also have a sustained impact for change to be realized at a systemic level.

Diversity, Equity, Inclusion, and Respect in Science

This work defines diversity, equity, inclusion, and respect (DEIR) as follows:

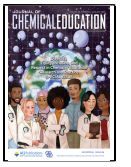
- Diversity: "differences in race, color, place of origin, religion, immigrant and newcomer status, ethnic origin, ability, sex, sexual orientation, gender identity, gender expression and age".[1]
- Equity: "the removal of systemic barriers and biases enabling all individuals to have equal opportunity to access and benefit from [a] program".[1]
- Inclusion: "the practice of ensuring that all individuals are valued and respected for their contributions and are equally supported".[1]
- Respect: a "regard for others as humans with something to contribute".[2]

This work also uses the term equity-deserving groups as a broad descriptor for the various historically marginalized groups in STEM. Terms like "equity-seeking" overlook the fact that marginalized groups have been seeking equity for decades, while "equity-deserving" explicates how all individuals have an inherent right to be treated equitably.[3]

Diversity in STEM is unrepresentative of the diversity in the general population.[4−6] A 2017 survey by the United States'

Special Issue: Diversity, Equity, Inclusion, and Respect in Chemistry Education Research and Practice

Received: April 28, 2021
Revised: August 24, 2021
Published: October 15, 2021



ACS Publications © 2021 American Chemical Society and Division of Chemical Education, Inc.

National Science Foundation (NSF) found that women, Black, Indigenous, and People of Color (BIPOC), and individuals with disabilities are underrepresented in the STEM workforce relative to the proportions of these groups within the general population.[5] In Canada, Aboriginal students are poorly represented in STEM relative to the general population.[7] A 2018 survey by the Royal Society of Chemistry in the United Kingdom found that students with disabilities in chemistry face unique challenges that do not arise in other subjects; women in chemistry face barriers to achieving leadership roles, scientific publishing, and awards, and the percentage of ethnic minority chemists drops significantly with increasing career stages.[8]

Causes for the lack of diversity observed in STEM are complex, multifaceted, and the focus of ongoing work on equity and inclusion.[2,9−11] One challenge in improving diversity is the absence of diversity itself—poor diversity contributes to a self-perpetuating cycle in which students of equity-deserving groups struggle to progress in STEM due to lack of role models.[12−15] A lack of role models has also been associated with imposter syndrome experienced by scientists from equity-deserving groups.[16] Another challenge is that support for scientific endeavors and access to a formal science education has historically excluded women, BIPOC, LGBTQ2S+ individuals, individuals with disabilities, and other equity-deserving groups.[17] This history has perpetuated a system in which certain individuals in society have greater access to education across generations, while students from equity-deserving groups are often first-generation students who lack role models within their personal, academic, and professional lives.[18] Who participates in science and engineering is also dependent on the state of DEIR at large—challenges faced by equity-deserving groups in STEM are a microcosm of the challenges faced by these groups in our broader society.[19−21] Broader historical and societal factors contribute to implicit biases and stereotypes that undermine equity-deserving groups in various facets of STEM (coursework, hiring, etc.).[22−28]

DEIR issues are not new; however, recent events have catalyzed a need for educators in STEM to listen to and identify actions that support equity-deserving groups in STEM. Much of this work has focused on undergraduate students, with less focused on graduate students. The limited focus on graduate students is concerning given recent research on the barriers faced by equity-deserving groups within graduate programs.[29,30] When discussing ways to improve DEIR in chemistry, a focus on graduate students will be imperative to enriching DEIR within the broader chemistry community. Graduate students are the future educators, researchers, and leaders in our shared chemistry community, and promoting DEIR within the graduate student population may bear positive downstream impacts on the state of DEIR in our broader chemistry community.

This article highlights how chemistry graduate student DEIR communities (CGDEIRCs) may drive sustained change related to DEIR practices and policies within the broader chemistry community. Since there is no formal definition for CGDEIRCs in the chemistry education literature, the authors—students engaged in CGDEIRCs—propose the following definition:

- CGDEIRCs are created, organized, and composed of graduate students in the chemical sciences.
- CGDEIRCs—including their goals, initiatives, events, etc.—focus on topics and issues related to enhancing DEIR within the chemical sciences.

CGDEIRCs have been prevalent at various academic institutions.[31] Many chemistry departments across the United States (US) have groups dedicated to women pursuing graduate studies and student chapters of the National Organization for the Professional Advancement of Black Chemists.[32−36] In Canada, the Canadians Working for Inclusivity in the Chemical Sciences, Engineering and Technology (CWIC) Network has been active since 2017: a national-level chemistry graduate student DEIR community. While CGDEIRCs are primarily targeted at the graduate student population, other community members (such as undergraduate students, faculty, and staff) are often encouraged to participate in and support these communities.

Drawing on the authors' experiences as students from four CGDEIRCs in North America—the CWIC Network, the University of Ottawa (uOttawa) CWIC Chapter, the University of Iowa DEIR Cohort, and the University of Minnesota-Twin Cities Graduate Student DEIR Community—this work will provide insight into:

- What chemistry graduate students may gain from participating in these communities.
- Examples of activities and initiatives these communities in the US and Canada have been engaged in.
- How these communities may afford a more diverse, equitable, and inclusive chemistry community, both in the present and in the future.
- How the chemistry community can continue to support these communities moving forward.

In addressing each of these items, this work contributes to the growing body of science and chemistry education literature, with a focus on the following items listed in the scope for this special issue:[2]

- Describing mentoring, teaching, and research methods for studying issues related to diversity, equity, inclusion, and respect.
- Sharing practices to reduce identity-related anxiety for learners and increase learner agency.
- Characterizing counter spaces, advocacy, and support structures in learning work environments.

Note on Authorship

Given the content of this manuscript, it is important to note that the authors of this manuscript serve as students and student leaders in voluntary chemistry graduate student DEIR communities at various North American institutions. All authors identify as individuals with historically marginalized identities in STEM and all have been driven to take part in CGDEIRCs because of their experiences with these identities in STEM. Due to the limited research literature on graduate student communities and how to enhance DEIR within chemistry graduate student populations, we situate this work within established theoretical and empirical work on learning communities, community cultural wealth, and cultural change, as well as our own anecdotal experiences as student members in CGDEIRCs. We acknowledge that while, these frameworks provide a rich body of literature to support our claims, they do not serve as a substitute for research on the impact(s) of chemistry graduate student DEIR communities.

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

CHEMISTRY GRADUATE STUDENT DEIR COMMUNITIES AS LEARNING COMMUNITIES

Chemistry graduate student DEIR communities (CGDEIRCs) are a student-type learning community (Supporting Information).[37,38] Student-type learning communities are communities that are specially designed for targeted groups, such as academically underprepared students, historically marginalized students, students with disabilities, etc.[39] In general, learning communities foster active learning over passive learning, cooperation over competition, community over isolation, and operationalize constructivist approaches to learning by providing opportunities for students to construct knowledge through interaction.[40]

CGDEIRCs provide a learning community for students to come together to discuss DEIR issues/topics and engage in a range of activities (Figure 1). In most cases, the students who organize and compose these communities are individuals who (a) identify with the goals of the community and (b) are equity-deserving individuals in STEM (women, BIPOC, LGBTQ2S+, individuals with disabilities, etc.).

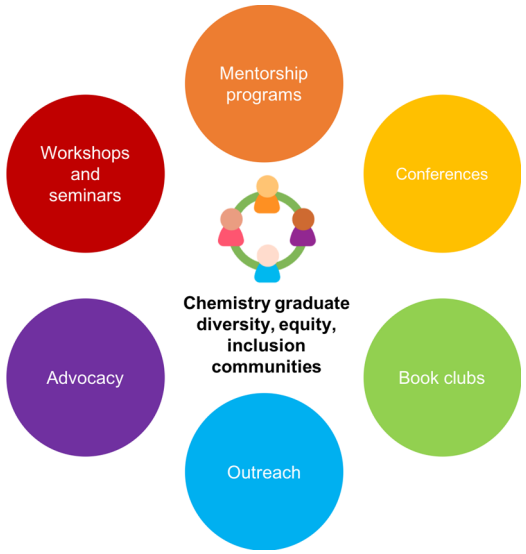


Figure 1. CGDEIRCs engage in a variety of initiatives to support students from equity-deserving groups and student learning on DEIR issues in chemistry.

CGDEIRCs aim to provide students with an increased capacity to reflect on their actions, thoughts, and behaviors in ways that contribute to a more equitable and inclusive chemistry community. For example, book clubs and seminar series organized by CGDEIRCs provide participants with opportunities to discuss specific DEIR issues, allowing for individuals to share their experiences and hear others' perspectives to construct a communal understanding of a particular DEIR topic, issue, etc. Co-construction of knowledge is key to learning communities, as learning communities are grounded in constructivist and sociocultural theories of learning.

Existing research on learning communities has been focused on the impacts of these types of communities on formal disciplinary contexts—for example, exploring how curricular learning communities impact undergraduate students' performance during a single program of study or impact student retention within a single institution.[41,42] Across various studies, learning communities have been found to positively impact students, with sources citing increased academic effort, openness to diversity, social tolerance, personal and interpersonal development, and stronger affinity to peers.[43−48] Undergraduate chemistry learning communities have been found to foster heightened interest and retention in science.[49] Learning communities have also been found to promote graduate student success, especially for students from equity-deserving groups in STEM.[50]

No research has directly investigated the impact(s) of student-led communities focused on DEIR. However, the available literature on learning communities suggests that CGDEIRCs can have positive outcomes for students who engage with them and their activities.

CHEMISTRY GRADUATE STUDENT DEIR COMMUNITIES AS SOURCES OF CULTURAL CAPITAL

By defining CGDEIRCs as learning communities, what might be the intended outcomes and opportunities that exist for students who take part in these communities? Answering this question requires considering the makeup of CGDEIRCs and the principal reasons for the existence of DEIR communities across various contexts (departments, professional societies, etc.).

The majority of CGDEIRCs are created, participated, and supported by equity-deserving groups in STEM who have identified a need for spaces to share experiences, learn from others, and find ways to improve DEIR in the broader chemistry community. For example, "women in STEM" groups are created, participated in, and supported by women who have identified a need for a space to share experiences, learn from each other, and establish ways to improve and advocate for the experience of being a woman in STEM within the chemical sciences. Drawing on the framework of *community cultural wealth*, CGDEIRCs provide students with positive outcomes centered on cultural wealth, providing them with skills and opportunities to attain historically exclusive forms of cultural capital.

Community cultural wealth (CCW) represents the array of knowledge, skills, abilities, and contacts utilized by equity-deserving communities to survive and resist macro- and microforms of oppression.[51] CCW emerges from critical race theory (CRT)[52] and was first proposed as a way to challenge deficit models of thinking in education for students of color. Research in science and engineering education has since applied the idea of CCW to other marginalized groups in STEM, such as women, transfer students, first-generation students, and students with disabilities.[53] CCW adopts an assets-based view of what students from marginalized communities bring to their education and work. Through a CCW lens, an educator or researcher seeks to identify strategies that equity-deserving students use to persevere through inequity and exclusion; this provides an avenue for which the unique experiences of equity-deserving students can be integrated within education to nurture and empower them. An assets-based approach contrasts against a deficit-based approach, which emphasizes what experiences students do not have, and then identifies ways to "supply" students with "accepted" skills and knowledge.

CCW can be developed within CGDEIRCs through activities that build capital. Activities can range from major events (conferences, workshops, seminars, etc.) or casual spaces for interactions between members of a community. The main types of capital and examples of initiatives within CGDEIRCs are summarized in Table 1. Many CGDEIRC initiatives are similar

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

Table 1. Summary of Types of Cultural Capital, Definitions, and Examples of Initiatives to Support Each Type

Type of Cultural Capital	Definitions[51]	General Examples of CGDEIRC Initiatives
Aspirational	"The ability to maintain hopes and dreams for the future, even in the face of real and perceived barriers"	Community-building activities and/or interactions with other community students that motivate an individual to persevere within their chemistry education/career
Linguistic	"Intellectual and social skills attained through communication experiences in more than one language and/or style"	Initiatives focused on students from shared, non-English linguistic backgrounds can provide students with opportunities to share experiences, build relationships, and navigate STEM institutions together
Familial	"Cultural knowledge nurtured among famil[ies] (kin) that carry a sense of community history, memory and cultural intuition. This form of cultural wealth engages a commitment to community well-being and expands the concept of family to include a [broader] understanding of kinship"	Spaces that provide individuals within the community to uncover and exchange shared experiences unique to equity-deserving individuals in STEM
Social	"Networks of people and community resources. These peer and other social contacts can provide both instrumental and emotional support to navigate through society's institutions"	Resources that help peers find/attain research awards, scholarships, etc.
Navigational	"Skills of maneuvering through social institutions. Historically, this infers the ability to maneuver through institutions not created with [students of equity-deserving groups] in mind"	Initiatives that support equity-deserving individuals' abilities to advocate for themselves in STEM environments and to properly handle discriminatory situations in their education and/or workplace
Resistant	"Knowledges and skills fostered through oppositional behavior that challenges inequality"	Providing students from equity-deserving groups a community in which their voices are fundamentally centered and creating spaces for students to come together and advocate for themselves and each other

to those discussed in a recent review on the CCW framework in science education research.[53] For example, connecting students with faculty members and advisors to discuss co-op/internship and research opportunities, advising students on professional development, and referrals to offices and on-campus services can support students' capital. Other work has found that students who share identities, challenges, and experiences create extended "families" through student organizations focused on community service and mentoring.[54−58] Though CCW has yet to be used as a framework in the chemistry education literature, its central ideas are reflected in chemistry education research on ways to support the retention of scientific identities in students from equity-deserving groups.[49,50,59,60]

As students gain wealth, this can facilitate academic/career success by providing students with access to historically inaccessible opportunities.[53] As an increasing number of equity-deserving individuals find success in chemistry due to CGDEIRC participation, issues such as poor diversity and inadequate visible role models in chemistry may be remedied in a sustained fashion.

Examples of the authors' different CGDEIRCs and activities are presented in Table 2, along with the different types of cultural capita they intend to build within participants. For readers interested in creating and/or supporting similar initiatives in their own communities, additional details related to each initiative are provided in the Supporting Information.

CHEMISTRY GRADUATE STUDENT DEIR COMMUNITIES AS CULTURAL CHANGE AGENTS

Cultural change is analogous to systemic change.[61,62] While various models for cultural change exist, this work relies on the four-frame model by Reinholz and Apkarian.[63,64] This model is employed herein because of its use in guiding STEM organizational change and its potential for investigating how CGDEIRCs can contribute to DEIR cultural change in chemistry. Reinholz and Apkarian's model is adapted from Bolman and Deal's 2008 model from literature on organizational management, and defines culture as a "historical and evolving set of structures and symbols and the resulting power relationships between people".[65] An organization's culture can be composed of various subcultures that spontaneously and continuously change over time. For example, the broader chemistry community has its own unique culture, but various subcultures (e.g., societies, departments, research groups, CGDEIRCs) can exist and align with the broader culture to varying degrees. Culture is inevitably and continuously changing—being aware of this and being able to identify goals for change can allow one to modify a culture in specific and meaningful ways. Change is achieved when one can identify how one or many structures, symbols, people, and power dynamics should change and then propose targeted initiatives that address these changes at various institutional levels.

Much of the previous work on achieving cultural change in STEM education has focused on how to increase uptake of evidence-based practices, such as active learning, in STEM departments. Identifying effective ways to enact change continues to be a challenge in education and educational research communities. One issue with change efforts is that most are guided by overly simplistic theories of change.[66−68] Systemic change related to DEIR has consistently been shown to be more complex than hosting one-off DEIR workshops and/or training sessions;[69,70] enacting and sustaining change requires a systemic approach, considering the chemistry community as a complex

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

Table 2. List of CGDEIRC Initiatives, along with Types of Capital They Seek to Develop[a]

Initiatives (details provided in the Supporting Information)	Type(s) of Capita Gained from Activities
(A) Leaders Overcoming Great Inequalities in Chemistry (LOGIC) Conference: Conference that hosts DEIR seminars, networking opportunities, workshops, flash talks, and poster sessions	Navigational: Creating a space for students to gain career advice and mentorship from those in more advanced positions in the chemical sciences
	Social and familial: Opportunities to contribute to working groups—in these groups, participants can share their thoughts on topics relating to mental health in graduate school, intersectionality in STEM, LGBTQ2S+ in STEM, transitioning from academia to industry, and others. The actionable suggestions from these group discussions are published on the CWIC Network Web site as a preliminary resource for other groups to engage in similar, sustained discussions and enact change.
	Aspirational: Opportunities for attendees to present their scientific research alongside their contributions to DEIR through flash talks and through opportunities to network with invited speakers and panelists in advanced career positions.
(A) My STEM Stories: Science communication workshop that focuses on providing diverse and unique examples of science-related career paths through a story-telling lens to improve visibility of equity-deserving groups in STEM education resources	Linguistic, navigational, and aspirational: Equipping participants with the skills to create videos describing their scientific work and journeys as an equity-deserving individual in STEM, allowing them to both identify what capital they have developed and relied on throughout their academic journeys, as well as generating aspirational capital within viewers
(B) DEIR-focused book club: Weekly book club focused on connecting students and starting conversations related to the challenges faced by equity-deserving groups both in STEM and in the general population	Social: Creating a space for conversations related to the challenges faced by equity-deserving groups both in STEM and in the general population
	Familial: Sharing personal/professional experiences related to topics of discussion
(B) Social media highlighting chemists from equity-deserving groups: social media campaign to highlight individuals from equity-deserving groups within the community	Aspirational: Amplifying the achievements of chemists from marginalized backgrounds to improve the visibility of role models for students from similar backgrounds
(B) Graduate student buddy program: Mentorship program in which junior graduate students are paired with upper-year mentors to ease transition into graduate school	Aspirational and navigational: Pairing new graduate students with an upper-year mentor to ease navigating the transition into to graduate school
	Social: Creates a strong network of future students, current students, and alumni through ongoing mentorship connections
(C) Underrepresented Students in STEM Symposium-Twin Cities: Local research symposium for BIPOC students to present their research and receive career mentorship	Aspirational: Inviting students to present their work at UMN-TC to afford them opportunity to improve their academic/career standing
	Social: Pairing students with graduate student mentors to assist them in creating posters and presentations; empowering mentees to reach out for further resources regarding undergraduate research experiences (URE), internship positions, and graduate school applications
(C) Research Experience for Undergraduates Workshop: Workshop for BIPOC students to learn about research opportunities and meet with potential supervisors	Social, aspirational, and navigational: Inviting students to bring their CVs/resumes to meet with the coordinators to strengthen their application based on their program(s) of interest; providing students with met with professors and graduate students from different programs at UMN to discuss graduate school, REUs, and "a day in the life" in graduate school.
(D) Microaggressions TA Training: Annual student-led training to equip graduate teaching assistants with heightened sensitivity to identity and microaggressions within the classroom	Resistant: Creating a microaggressions reporting form; creating a microagressions TA training program to support students' abilities to identify and challenge systemic bias
(D) Colloquium series on DEIR graduate student leadership: Department-wide seminars for graduate students to present on their work in DEIR	Social and navigational: Professional development seminars and workshops focused on social justice themes (e.g., intersectionality, racism, collective trauma, etc.) that connect students with peers involved in DEIR work
(D) Identity awareness week: A week of DEIR-based spaces for discussion, support, and networking centered around marginalized groups in STEM	Familial: Provides a space for discussion, support, and networking centered around four highly marginalized groups in STEM: 1) LGBTQ2S+ scientists, 2) women in science, 3) scientists with visible/invisible disabilities and mental health challenges, and 4) international scientists

[a]Details about specific initiatives are in the Supporting Information. (A) CWIC Network, (B) uOttawa CWIC Chapter, (C) University of Minnesota-Twin Cities (UMN-TC) Graduate Student DEIR Community, and (D) University of Iowa Graduate Student DEIR Cohort.

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

social system in which sustained efforts from various entry points will be most effective.[63,71,72]

CGDEIRCs and their student members contribute to cultural change in both active (e.g., championing, advocating for, and modeling DEIR practices throughout their careers and within their communities) and passive ways (e.g., chemistry/STEM career progress, becoming a role model for students from similar backgrounds). How CGDEIRCs and their student members might shape cultural structures, symbols, people, and power dynamics will be discussed in the following subsections.

Structures

Structures are defined as the "roles, responsibilities, practices, and incentives that organize how people [within a community] interact".[63] Structures might include what formal positions exist (e.g., faculty vs staff), committee structures, etc. Structures define the possible roles and their expectations in a community and communicate what is considered acceptable and important.

Structural change is attained when there is a focus on broad, positive, achievable outcomes. A focus on outcomes allows a community to engage in productive conversation, rather than short-term, inflexible solutions. For example, the CWIC Network provides sustained space for students and scientists from equity-deserving groups with consistent visibility and rich interactions that build cultural wealth over time.

CGDEIRCs also challenge the explicit structure of STEM by providing spaces where DEIR issues, conversations, and equity-deserving voices are explicitly centered. Networks like CWIC and community spaces like the uOttawa CWIC chapter and UI Graduate Student DEIR cohort provide students with frequent opportunities to engage in DEIR discussions through book clubs, seminars, and conferences. Students in these communities regularly interact and learn from each other, share perspectives, issues, and solutions, and work together toward shared community success and communicate DEIR values to their broader community.

CGDEIRCs can also provide students with opportunities and outcomes historically inaccessible to them. For example, the CWIC Network's LOGIC Conference and Underrepresented Students in STEM Symposium-Twin Cities (USSS-TC) provide a space for equity-deserving students to present their research, allowing them to learn from others who share experiences and build professional skills together (e.g., public speaking, networking). Awards in CGDEIRCs are often focused on supporting students of equity-deserving groups, and thus, these also serve as ways to build professional capital. Initiatives may also be inspired by personal and professional challenges faced by students, allowing them to share and validate experiences and work toward solutions together.

Symbols

Symbols constitute the cultural artifacts, language, knowledge, myths, values, and vision that community members use to guide their reasoning.[63] Symbols can be considered the underlying ways of thinking that give meaning to the structures in a department.

CGDEIRCs enhance how DEIR is valued in chemistry communities. Activities organized by CGDEIRCs actively and explicitly center DEIR. While these spaces often focus on specific topics related to DEIR (e.g., microaggressions, implicit bias), they broadly serve as a platform for students to reflect on their actions, thoughts, and behaviors, which can lead to symbolic shifts related to DEIR (e.g., learning how to behave in more inclusive ways). For example, initiatives like the micro-aggressions training (University of Iowa) and the DEIR book club (uOttawa) provide students with opportunities to reflect on the DEIR principles and how to better enact these principles in their thoughts and behaviors. CGDEIRCs can also contribute to symbolic shifts beyond their immediate membership. The presence and scope of the CWIC Network in Canada has shaped how national societies, such as the Canadian Society of Chemistry, consider DEIR in communications, conference programming, etc. CGDEICs can also provide students to work with existing staff/faculty DEIR communities, expanding the scope and impact of symbolic shifts. For example, the uOttawa CWIC Chapter and UI Graduate Student DEIR Cohort have worked with faculty members and existing faculty DEIR committees to facilitate dialogue between students and faculty, which has resulted in coordinated initiatives and messaging to enhance DEIR at various levels, though data is needed to adequately measure the impacts of these coordinated initiatives/messaging on the broader community.

People and Power

A community is composed of individuals, each with their own goals, agency, needs, and identities, and interactions between individuals are mediated by power dynamics within the community.[63] Considering the people in a culture and the power dynamics between them is essential for challenging established norms related to equity and inclusion. Power dynamics can be based on aspects of identity (race, gender, sexual orientation, ability, etc.), such as how historical advantages have been afforded to certain groups at the expense of other groups.[73] This frame of the model is particularly relevant to CGDEIRCs as spaces that focus on DEIR and are primarily composed of members of equity-deserving groups in chemistry.

Changing the culture of DEIR in chemistry demands considering what power dynamics exist within a culture and identifying ways to reshape these dynamics to become more equitable and inclusive. Considering the narratives that have historically dominated chemistry, CGDEIRCs challenge existing power dynamics through (a) their existence as a space inclusive and supportive of equity-deserving groups and (b) activities that provide equity-deserving students with opportunities and activities to help them persist in the broader community. In "leveling the playing field", CGDEIRCs also contribute to sustained shifts in power by supporting the retention of students from equity-deserving groups, who by persisting, act as active and passive change agents within the community (e.g., serving on institutional committees and modeling/practicing/sharing DEIR values in these spaces, advancing in scientific careers and becoming role models for students from similar backgrounds).[9,74]

CGDEIRCs can also help mitigate traditional student-faculty power dynamics in chemistry. Within an institution, those with the resources to directly initiate immediate change are often those who hold power within the institution. CGDEIRCs provide opportunities for students to effectively engage with faculty/administrators in conversations about DEIR by uniting students around a common goal/message. For example, recent events at the University of Ottawa motivated the uOttawa CWIC Chapter to organize a departmental letter from students pushing for better consideration of DEIR in decision-making. By leveraging the collective voice of a CGDEIRC, student voices can be better positioned in policy-level decisions related to DEIR. This can lead to sustained participation; for example, at

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

the Universities of Ottawa and Iowa, students from CGDEIRCs are actively invited to faculty DEIR committees. The effectiveness of such actions will depend on how student-faculty power dynamics are considered within a community. The most impactful interactions between students and faculty have emerged when both parties explicitly express a desire to collaborate with, listen to, and learn from each other's experiences.

OPPORTUNITIES FOR FACULTY SUPPORT

As noted earlier, the authors of this work are students who have been engaged in CGDEIRCs. Acknowledging that the principal audience of the *Journal of Chemical Education* are faculty members and educators in the chemical sciences, this work concludes by sharing challenges that CGDEIRCs face that may be mitigated through with faculty support (list in the Supporting Information). CGDEIRCs represent a bottom-up approach to facilitating change, but this must be paired with a top-down approach to realize significant change.[72,75]

First, CGDEIRCs are, by definition, student-led. A grassroots community has distinct advantages, such as the flexibility to discuss topics that are relevant and interesting to students in that community. However, challenges may include (1) a lack of dedicated and consistent funding for initiatives; (2) variable engagement depending on students' research commitments and expectations; and (3) as discussed in "People and Power" limited access to policy discussions that may directly impact students within an institution. Faculty can lessen these challenges by (1) sharing/creating funding opportunities with students from these communities; (2) explicitly communicating the value of participating in CGDEIRC initiatives within research groups, lectures, etc.; and (3) treating students as partners on relevant committees to engage in active dialogue.

Second, CGDEIRCs are primarily participated by students from equity-deserving groups. Though this has distinct advantages by appealing to the goals and interests of students, this can come at the risk of overburdening equity-deserving students with solving complex systemic issues. If the community values CGDEIRCs, their efforts must be explicitly supported and shared by faculty with students, researchers, and colleagues. For example, faculty could share CGDEIRC initiatives and resources with students (in lectures, email communications, etc.) to champion DEIR in science, participate in or leading faculty-level DEIR initiatives alongside students, and prioritize space/time for DEIR discussions (e.g., DEIR-focused group meetings). In the long term, this may alleviate the longstanding issue of devaluing service work, which has been shown to specifically hinder the advancement of academics from equity-deserving groups.[76,77] Such efforts may also propagate a broader appreciation of DEIR in the chemistry community, minimizing misinformed attitudes toward DEIR work.[78]

Third, research is needed on the impact(s) of CGDEIRCs. Though this work draws on prior literature on learning communities, community cultural wealth, and cultural change to support the claims advanced in this article, this is not a substitute for data. Potential investigations may include qualitative studies (interviews, focus groups, etc.) with students who actively participate in CGDEIRCs to investigate how and what types of community cultural wealth/capital are developed (if it all) within these communities and through what actions and initiatives. Implications can provide graduate students and faculty with better understanding of how and why specific initiatives support community members.

CONCLUSION

Change efforts and educational initiatives related to improving DEIR in chemistry are ongoing and increasing. This contribution highlights how graduate student DEIR communities may contribute to cultural change related to DEIR by providing students with various activities and opportunities to build community cultural wealth. With respect to items from the special issue scope:

- CGDEIRCs provide students with mentoring, teaching, and research opportunities for studying issues related to diversity, equity, inclusion, and respect through communities and initiatives focused on DEIR issues in STEM.
- CGDEIRCs reduce identity-related anxiety for learners and increase learner agency by providing opportunities for students from equity-deserving groups to gain cultural capital, advocate for themselves and their identities, and gain professional skills and experiences within peer networks.
- CGDEIRCS represent a counter space and support structures for students to advocate for equity and inclusion issues in STEM within the broader chemistry community

From the authors' experiences in these communities as students, a list of CGDEIRC activities across North America was created (Table 2, with details in the Supporting Information) with the hope that readers interested in starting similar communities and events can find ways to do so in their own setting. We also propose ways faculty can support these communities to lessen the burden on students from equity-deserving groups. Given the lack of empirical research on graduate student learning communities and chemistry DEIR learning outcomes, we situate this work in a breadth of literature focused on learning communities, community cultural wealth, and cultural change in hopes that future research in the field might use these frameworks to investigate the impacts of CGDEIRCs. Driving cultural shifts toward a more diverse, equitable and inclusive chemistry community will require careful consideration of how to promote these values among all learners at all levels, and CGDEIRCs provide a foundation for educators and researchers to investigate how to improve DEIR at the graduate student level.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00485.

Learning community definitions, list of CGDEIRC initiatives, and list of opportunities for faculty support (PDF)

AUTHOR INFORMATION

Corresponding Author

Jacky M. Deng − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada;* orcid.org/0000-0002-2633-6624; Email: [redacted]

Authors

Leah E. McMunn − *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada;* orcid.org/0000-0002-0293-4080

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

Meagan S. Oakley − *Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada;* orcid.org/0000-0001-5072-7572
Hoang T. Dang − *Department of Chemistry, University of Iowa, Iowa City, Iowa 52242, United States*
Rebeca S. Rodriguez − *Department of Chemistry, University of Minnesota Twin Cities, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0002-8994-554X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.1c00485

Author Contributions

⊥L.E.M., M.S.O., H.T.D., and R.S.R. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.S.O. would like to thank the Natural Science and Engineering Research Council for funding for My STEM Stories (NSERC Award 556068-2020) and the University of Toronto Department of Chemistry for funding the CWIC Network and CWIC Chapter events. R.S.R. would like to thank the local MN ACS Chapter for their funding of *Spooky Science* and the UMN Medical School for the SACNAS funding in the REU Workshop. L.E.M. would like to thank the CWIC Network and Campus Vibez uOttawa (the coordinating body for clubs at the University of Ottawa) for their funding of the DEIR library and book club at the University of Ottawa. H.T.D. would like to thank the UI Department of Chemistry for their funding of the colloquium series on DEIR graduate student leadership. J.M.D. would like to thank Dr. Alison B. Flynn for providing feedback on an early version of this manuscript. Lastly, all authors would like to thank the students and faculty who have supported DEIR work at their respective institutions—many of the initiatives and communities described in this work would not be possible without their collective support.

■ REFERENCES

(1) Government of Canada. Best Practices in Equity, Diversity and Inclusion in Research. https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/edi-eng.aspx (accessed April 21, 2021).
(2) Winfield, L. L.; Wilson-Kennedy, Z. S.; Payton-Stewart, F.; Nielson, J.; Kimble-Hill, A. C.; Arriaga, E. A. Journal of Chemical Education Call for Papers: Special Issue on Diversity, Equity, Inclusion, and Respect in Chemistry Education Research and Practice. *J. Chem. Educ.* **2020**, *97* (11), 3915−3918.
(3) Smith, T. Shifting from equity-seeking to equity-deserving. https://wordpress.kpu.ca/antiracism/2020/11/20/shifting-from-equity-seeking-to-equity-deserving/ (accessed 2021-08-15).
(4) Whitcomb, K. M.; Singh, C. Not All Disadvantages Are Equal: Racial/ethnic Minority Students Have Largest Disadvantage of All Demographic Groups in Both STEM and Non-STEM GPA. *arXiv* **2020**, *15260*, 1−12.
(5) National Science Foundation. *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017 Digest Special Report NSF 17−310*; 2017.
(6) National Science Foundation. Doctoral Scientists and Engineers in Academia. https://www.nsf.gov/statistics/2016/nsb20161/#/report/chapter-5/doctoral-scientists-and-engineers-in-academia/federal-support-of-doctoral-researchers-in-academia (accessed 2020-11-15).
(7) Etkin, N.; Armour, M.-A.; Franz-Odendaal, T.; Rayner-Canham, G.; Rayner-Canham, M.; Al-Abadleh, H. A.; Jones, K. *Equity and Diversity in Chemistry*; MacMillan Learning Curriculum Solutions: Plymouth, MI, 2016.
(8) The Royal Society of Chemistry. *Diversity Landscape of the Chemical Sciences*; 2018.
(9) Posselt, J. R. *Equity in Science: Representation, Culture, and the Dynamics of Change in Graduate Education*; Stanford University Press: Stanford, CA, 2020.
(10) Sanford, M. Equity and Inclusion in the Chemical Sciences Requires Actions Not Just Words. *J. Am. Chem. Soc.* **2020**, *142*, 11317.
(11) Tilghman, B. S.; Alberts, B.; Colón-Ramos, D.; Dzirasa, K.; Kimble, J.; Varmus, H. Concrete Steps to Diversify the Scientific Workforce. *Science* **2021**, *372*, 133−136.
(12) González-Pérez, S.; Mateos de Cabo, R.; Sáinz, M. Girls in STEM: Is It a Female Role-Model Thing? *Front. Psychol.* **2020**, *11*, 2204.
(13) Breda, T.; Grenet, J.; Monnet, M.; van Effenterre, C. Can Female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools. *Hal* **2018**, 18−19.
(14) Beasley, M. A.; Fischer, M. J. Why They Leave: The Impact of Stereotype Threat on the Attrition of Women and Minorities from Science, Math and Engineering Majors. *Soc. Psychol. Educ.* **2012**, *15* (4), 427−448.
(15) Cheryan, S.; Master, A.; Meltzoff, A. N. Cultural Stereotypes as Gatekeepers: Increasing Girls' Interest in Computer Science and Engineering by Diversifying Stereotypes. *Front. Psychol.* **2015**, *6* (FEB), 1−8.
(16) Funk, C.; Parker, K. Women and Men in STEM Often at Odds Over Workplace Equity. *Pew Res. Cent.* **2018**, 1−19.
(17) Woolston, C. White Men Still Dominate in UK Academic Science. *Nature* **2020**, *579*, 622.
(18) Kredell, M. Equity Runs Through It. *USC Rossier* **2017**, 20−23.
(19) Fuentes, E.; Liou, D. Language, Culture, and Access. In *Unfinished Business: Closing the Racial Achievement Gap in Our Schools*; Nogueira, P. A., Wing, J. Y., Eds.; Jossey-Bass: San Francisco, CA, 2008; pp 52−68.
(20) Flynn, D. T. STEM Field Persistence: The Impact of Engagement on Postsecondary STEM Persistence for Underrepresented Minority Students. *J. Educ. Issues* **2016**, *2* (1), 185.
(21) Lee, V. E.; Burkam, D. T. *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*; Economic Policy Institute, 2002.
(22) Isgro, K.; Castaneda, M. Mothers in U.S. Academia: Insights from Lived Experiences. *Womens. Stud. Int. Forum* **2015**, *53*, 174−181.
(23) Barrett, L. F.; Bliss-Moreau, E. She's Emotional. He's Having a Bad Day: Attributional Explanations for Emotion Stereotypes. *Emotion* **2009**, *9* (5), 649−658.
(24) Reuben, E.; Sapienza, P.; Zingales, L. How Stereotypes Impair Women's Careers in Science. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (12), 4403−4408.
(25) Sarsons, H. *Gender Differences in Recognition for Group Work*; Harvard University, **2015**.
(26) O'Brien, L. T.; Blodorn, A.; Adams, G.; Garcia, D. M.; Hammer, E. Ethnic Variation in Gender-STEM Stereotypes and STEM Participant: An Intersectional Approach. *Cult. Divers. Ethn. Minor. Psychol.* **2015**, *21* (2), 169−180.
(27) Yoder, J. B.; Mattheis, A. Queer in STEM: Workplace Experiences Reported in a National Survey of LGBTQA Individuals in Science, Technology, Engineering, and Mathematics Careers. *J. Homosex.* **2016**, *63* (1), 1−27.
(28) Moss-Racusin, C. A.; Molenda, A. K.; Cramer, C. R. Can Evidence Impact Attitudes? Public Reactions to Evidence of Gender Bias in STEM Fields. *Psychol. Women Q.* **2015**, *39* (2), 194−209.
(29) Jaquette, O.; Curs, B. R.; Posselt, J. R. Tuition Rich, Mission Poor: Nonresident Enrollment Growth and the Socioeconomic and Racial Composition of Public Research Universities. *J. Higher Educ.* **2016**, *87* (5), 635−673.
(30) Posselt, J. R.; Grodsky, E. Graduate Education and Social Stratification. *Annu. Rev. Sociol.* **2017**, *43*, 353−378.
(31) Bonetta, L. Affinity Groups for Diversity. *Science*. **2008**.

(32) Santos-Díaz, S.; Towns, M. H. An All-Female Graduate Student Organization Participating in Chemistry Outreach: A Case Study Characterizing Leadership in the Community of Practice. *Chem. Educ. Res. Pract.* **2021**, *22*, 532.
(33) Harvard Women in Chemistry. https://hwpi.harvard.edu/womeninchemistry/home (accessed 2021-04-01).
(34) *Women+ in Chemistry*. https://women.chem.upenn.edu/#:~:text=Welcome (accessed 2021-04-01).
(35) Nguyen, T. Supporting women in STEM through safe spaces, networking, and empowerment. https://chemistry.princeton.edu/news/fric-supporting-women-stem-through-safe-spaces-networking-and-empowerment (accessed 2021-04-01).
(36) NOBCChE Student Chapters. *National Organization for the Professional Advancement of Black Chemists and Chemical Engineers*. https://www.nobcche.org/student-chapters2 (accessed 2021-04-10).
(37) West, R. E.; Williams, G. S. I Don't Think That Word Means What You Think It Means": A Proposed Framework for Defining Learning Communities. *Educ. Technol. Res. Dev.* **2017**, *65* (6), 1569−1582.
(38) Zhao, C.-M.; Kuh, G. D. Adding Value: Learning Communities and Student Engagement. *Res. High. Educ.* **2004**, *45* (2), 115−138.
(39) Lenning, O.; Ebbers, L. *The Powerful Potential of Learning Communities: Improving Education for the Future. ASHE-ERIC Higher Education Report, Vol. 26, No. 6*; ERIC Publications, 1999.
(40) Cross, K. P. Why Learning Communities? Why Now? *About Campus* **1998**, *3*, 4−11.
(41) Shochet, R.; Fleming, A.; Wagner, J.; Colbert-Getz, J.; Bhutiani, M.; Moynahan, K.; Keeley, M. Defining Learning Communities in Undergraduate Medical Education: A National Study. *J. Med. Educ. Curric. Dev.* **2019**, *6*, 238212051982791.
(42) Baker, S.; Pomerantz, N. Impact of Learning Communities on Retention at a Metropolitan University. *J. Coll. Student Retent. Res. Theory Pract.* **2000**, *2* (2), 115−126.
(43) Johnson, D.; Johnson, R.; Smith, K. Cooperative Learning Center Directors. *Annu. Rep. Educ. Psychol. Japan* **1998**, 1−29.
(44) Whitt, E. J.; Edison, M. I.; Pascarella, E. T.; Terenzini, P. T.; Nora, A. Influences on Students' Openness to Diversity and Challenge in the Second and Third Years of College. *J. Higher Educ.* **2001**, *72* (2), 172−204.
(45) Slavin, R. E. When Does Cooperative Learning Increase Student Achievement? *Psychol. Bull.* **1983**, *94* (3), 429−445.
(46) Hurtado, S. S.; Gonyea, R. M.; Graham, P. A.; Fosnacht, K. The Relationship Between Residential Learning Communities and Student Engagement. *American College Personnel Association* **2019**, *8* (1), 1−25.
(47) Pike, G. R. The Effects of Residential Learning Communities and Traditional Residential Living Arrangements on Educational Gains during the First Year of College. *J. Coll. Stud. Dev.* **1999**, *40* (3), 269−284.
(48) Tinto, V. Colleges as Communities: Taking Research on Student Persistence Seriously. *Rev. High. Educ.* **1998**, *21* (2), 167−177.
(49) deProphetis Driscoll, W.; Gelabert, M.; Richardson, N. Efficacy of Using Learning Communities To Improve Core Chemistry Education and Increase Student Interest and Retention in Chemistry. *J. Chem. Educ.* **2010**, *87* (1), 49−53.
(50) Wilson, Z. S.; McGuire, S. Y.; Limbach, P. A.; Doyle, M. P.; Marzilli, L. G.; Warner, I. M. Diversifying Science, Technology, Engineering, and Mathematics (STEM): An Inquiry into Successful Approaches in Chemistry. *J. Chem. Educ.* **2014**, *91* (11), 1860−1866.
(51) Yosso, T. J. Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth. *Race Ethn. Educ.* **2005**, *8* (1), 69−91.
(52) Delgado, R.; Stefancic, J. *Critical Race Theory*; New York University Press, 2015.
(53) Denton, M.; Borrego, M.; Boklage, A. Community Cultural Wealth in Science, Technology, Engineering, and Mathematics Education: A Systematic Review. *J. Eng. Educ.* **2020**, *109* (3), 556−580.
(54) Braun, D. C.; Gormally, C.; Clark, M. D. The Deaf Mentoring Survey: A Community Cultural Wealth Framework for Measuring Mentoring Effectiveness with Underrepresented Students. *CBE Life Sci. Educ.* **2017**, *16* (1), 1−14.
(55) Martin, J. P.; Newton, S. S. Uncovering Forms of Wealth and Capital Using Asset Frameworks in Engineering Education. *ASEE Annu. Conf. Expo. Conf. Proc.* **2016**, 16387.
(56) Foxx, K. A. A Cultural Wealth Perspective on Success and Persistence in Engineering for Black and Latino Students. *Exploring Interactions with Faculty and Other Institutional Agents*; University of North Carolina, 2014.
(57) Revelo Alonso, R. A. *Engineering familia: The Role of a Professional Organization in the Development of Engineering Identities of Latina/o Undergraduates*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, 2015.
(58) Samuelson, C. C.; Litzler, E. Community Cultural Wealth: An Assets-Based Approach to Persistence of Engineering Students of Color. *J. Eng. Educ.* **2016**, *105* (1), 93−117.
(59) Miller-Friedmann, J.; Childs, A.; Hillier, J. Approaching Gender Equity in Academic Chemistry: Lessons Learned from Successful Female Chemists in the UK. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 24−41.
(60) Fink, A.; Frey, R. F.; Solomon, E. D. Belonging in General Chemistry Predicts First-Year Undergraduates' Performance and Attrition. *Chem. Educ. Res. Pract.* **2020**, *21* (4), 1042−1062.
(61) Schein, E. H. *Organizational Culture and Leadership*; Josey-Bass: San Francisco, CA, 2010.
(62) Schein, E. H. Kurt Lewin's Change Theory in the Field and in the Classroom: Notes toward a Model of Managed Learning. *Syst. Pract.* **1996**, *9* (1), 27−47.
(63) Reinholz, D. L.; Apkarian, N. Four Frames for Systemic Change in STEM Departments. *Int. J. STEM Educ.* **2018**, *5* (1), 3.
(64) Reinholz, D. L.; Ngai, C.; Quan, G.; Pilgrim, M. E.; Corbo, J. C.; Finkelstein, N. Fostering Sustainable Improvements in Science Education: An Analysis through Four Frames. *Sci. Educ.* **2019**, *103* (5), 1125−1150.
(65) Bolman, L. G.; Deal, T. E. *Reframing Organizations: Artistry, Choice, and Leadership*; Josey-Bass: San Francisco, CA, 2008.
(66) Apkarian, N.; Kirin, D. Active Learning in Undergraduate Precalculus and Single Variable Calculus. In *Proceedings of the 19th Annual Conference on Research in Undergraduate Mathematics Educationnnual conference on research in undergraduate mathematics education*; Fukawa-Connelly, T., Infante, N., Wawro, M., Brown, S., Eds.; Pittsburgh, PA, 2016; pp 512−514.
(67) Borrego, M.; Henderson, C. Increasing the Use of Evidence-Based Teaching in STEM Higher Education: A Comparison of Eight Change Strategies. *J. Eng. Educ.* **2014**, *103* (2), 220−252.
(68) Henderson, C.; Beach, A.; Finkelstein, N. Facilitating Change in Undergraduate STEM Instructional Practices: An Analytic Review of the Literature. *J. Res. Sci. Teach.* **2011**, *48* (8), 952−984.
(69) Chang, E. H.; Milkman, K. L.; Gromet, D. M.; Rebele, R. W.; Massey, C.; Duckworth, A. L.; Grant, A. M. The Mixed Effects of Online Diversity Training. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (16), 7778−7783.
(70) Pruitt, A.-S.; Brinkworth, C.; Young, J.; Aponte, K. L. 5 Things We Learned About Creating a Successful Workplace Diversity Program. *Harvard Business Review*, **2018**
(71) Kezar, A. What Is the Best Way to Achieve Broader Reach of Improved Practices in Higher Education? *Innov. High. Educ.* **2011**, *36* (4), 235−247.
(72) Lawrie, G.; Marquis, E.; Fuller, E.; Newman, T.; Qiu, M.; Nomikoudis, M.; Roelofs, F.; van Dam, L. Moving Towards Inclusive Learning and Teaching: A Synthesis of Recent Literature. *Teach. Learn. Inq. ISSOTL J.* **2017**, *5* (1), 9−21.
(73) Smith, M. S.; Bray, N. U15 Leadership Remains Largely White and Male Despite 33 Years of Equity Initiatives. https://uofaawa.wordpress.com/2019/06/20/u15-leadership-remains-largely-white-and-male-despite-33-years-of-equity-initiatives/ (accessed 2020-10-14).
(74) Nielsen, W.; Alegria, S.; Börjeson, L.; Falk-Krzesinski, H. J.; Joshi, A.; Leahey, E.; Smith-Doerr, L.; Woolley, A. W.; et al. Gender Diversity

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

Leads to Better Science. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (13), E2796.
(75) Marquis, E.; Jung, B.; Fudge-Schormans, A.; Vajoczki, S.; Wilton, R.; Baptiste, S.; Joshi, A. Creating, Resisting or Neglecting Change: Exploring the Complexities of Accessible Education for Students with Disabilities. *Can. J. Scholarsh. Teach. Learn.* **2012**, *3* (2), 1−21.
(76) Wijesingha, R.; Ramos, H. Human Capital or Cultural Taxation: What Accounts for Differences in Tenure and Promotion of Racialized and Female Faculty? *Can. J. High. Educ.* **2017**, *47* (3), 54−75.
(77) Jimenez, M. F.; Laverty, T. M.; Bombaci, S. P.; Wilkins, K.; Bennett, D. E.; Pejchar, L. Underrepresented Faculty Play a Disproportionate Role in Advancing Diversity and Inclusion. *Nat. Ecol. Evol.* **2019**, *3* (7), 1030−1033.
(78) Impellizzeri, S.; Coe, I. R. The Complex Chemistry of Diversity and Inclusion: A 30-Year Synthesis. *Can. J. Chem.* **2021**, *99* (8), 653−660.

https://doi.org/10.1021/acs.jchemed.1c00485
J. Chem. Educ. 2022, 99, 373−382

Chapter 10: Analysis of financial challenges faced by graduate students in Canada

10.1 Introduction to the research article presented in this chapter

Chapter 10 describes a national survey study on the landscape of graduate student finances across Canada, titled the "National Graduate Student Finance Survey". The project was led by the Ottawa Science Policy Network (OSPN), a student-led science policy organization the University of Ottawa, of which I served as EDI Coordinator. We aimed to capture a snapshot of graduate student finances, including an in-depth analysis of stipends, scholarships, debt, tuition, and living expenses. We received responses from 1305 graduate students across Canada.

In my role, I analyzed demographic data collected from the survey (*e.g.*, race/ethnicity, gender, disability), connecting this to data on respondents' finances and experiences. I also wrote and edited related sections of OSPN's survey report, which served as the basis for the manuscript shared in this chapter.

Thank you to OSPN for giving me the opportunity to contribute to this project. In particular I want to recognize and thank Sarah Laframboise, whose leadership and advocacy continues to inspire me. Through this project, I was able to participate in important conversations about research funding with researchers from across Canada, as well as Members of Parliament in the Canadian House of Commons. These conversations gave me the incredible opportunity to see how science and politics interact and intersect, forever shaping how I approach science communication and outreach.

10.2 Copyright

Reprinted from "Laframboise, S. J.; Bailey, T.; Dang, A.-T.; Rose, M.; Zhou, Z.; Berg, M. D.; Holland, S.; Abdul, S. A.; O'Connor, K.; El-Sahli, S.; Boucher, D. M.; Fairman, G.; **Deng, J. M**.; Shaw, K.; Noblett, N.; D'Addario, A.; Empey, M.; Sinclair, K. Analysis of Financial Challenges

Faced by Graduate Students in Canada. *Biochem. Cell Biol.* **2023**, *101* (4), 326–360. doi.org/10.1139/bcb-2023-0021" with permission from Canadian Science Publishing.

Biochemistry and Cell Biology

OPEN ACCESS | Article

Analysis of financial challenges faced by graduate students in Canada

Sarah Jane Laframboise[a,b], **Thomas Bailey**[a,b], **Anh-Thu Dang**[a,b], **Mercedes Rose**[a,b], **Zier Zhou**[a,b], **Matthew D. Berg**[c], **Stephen Holland**[a,b], **Sami Aftab Abdul**[a,b], **Kaela O'Connor**[a,b], **Sara El-Sahli**[a,b], **Dominique M. Boucher**[a,b], **Garrett Fairman**[a,b], **Jacky Deng**[a,b], **Katherine Shaw**[a,b], **Nathaniel Noblett**[a,b], **Alexa D'Addario**[a,b], **Madelaine Empey**[a,b], **and Keaton Sinclair**[a,b]

[a]University of Ottawa, Ottawa, ON, Canada; [b]Ottawa Science Policy Network, University of Ottawa, Ottawa, ON, Canada; [c]Canadian Society of Molecular Biosciences, Ottawa, ON, Canada

Corresponding author: **Sarah Laframboise** (email: [redacted])

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Abstract

Graduate students are vital to the creation of research and innovation in Canada. The National Graduate Student Finance Survey was launched in 2021 by the Ottawa Science Policy Network to investigate the financial realities of Canadian graduate students. Closing in April 2022, the survey received 1305 responses from graduate students representing various geographical locations, years of study, fields of education, and demographic backgrounds. The results capture a snapshot into graduate student finances, including an in-depth analysis of stipends, scholarships, debt, tuition, and living expenses. In its entirety, we found that the majority of graduate students are facing serious financial concerns. This is largely due to stagnant funding for students both from federal and provincial granting agencies and from within their institutions. This reality is even worse for international students, members of historically underrepresented communities, and those with dependents, all of whom experience additional challenges that impact their financial security. Based on our findings, we propose several recommendations to the Tri-Council agencies (Natural Sciences and Engineering Research Council, Social Science and Humanities Research Council, and Canadian Institute for Health Research) and academic institutions to strengthen graduate student finances and help sustain the future of research in Canada.

Key words: Graduate Students, Canadian Research, Stipends, Funding, Scholarships, Tri-Council

Introduction

Graduate students play a large role in the research and innovation ecosystem in Canada. These students have already completed a bachelor's degree and are enrolled in a master's or doctoral degree. Most graduate degrees include full-time, independent research in collaboration with a Principal Investigator or supervisor, who is a faculty member at a university. This work requires laborious training and creates the foundation of research in Canada. Research has shown that one-third of publications feature PhD students as either authors or co-authors (Lariviére 2012), exemplifying the crucial role that graduate students play in research and academic publications.

Since graduate students are often enrolled in full-time studies, many receive a stipend intended to offset living costs so that they can focus on their studies. The source of this funding for graduate students can be complex. In Canada, funding comes from two main sources: (1) directly through scholarships, such as those provided by the three federal granting agencies, known as the Tri-Council (Natural Sciences and Engineering Research Council (NSERC), Social Science and Humanities Research Council (SSHRC), and the Canadian Institute for Health Research (CIHR)) or (2) indirectly through their supervisor's research grants and(or) department funding. As such, graduate student pay is often referred to as a stipend, since it is often a fixed pay, independent from work performed. Stipends can vary among graduate students and depend on many factors. For example, some students may be required to complete teaching assistantships (TAs) in which they assist a professor with instructional responsibilities or research assistantships (RAs) in which they assist a professor or instructor with additional research activities such as performance of clerical, laboratory, or technical tasks as part of their stipend.

The value of graduate student stipends can be set by universities, faculties, departments, and supervisors. It appears that, in the absence of any other government guidance, federal Tri-Council scholarship values have set the precedent for stipend values. However, it is important to acknowledge that these scholarships have not changed in value since 2003. Additionally, there has been little increase in the number of available Canadian scholarships over the years. A 2018 report from the Science and Policy Exchange (SPE) revealed that 91% of graduate students were largely displeased with

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

the number of Tri-Council scholarships available for graduate students (Science and Policy Exchange 2020). Furthermore, 79% replied that they would like to see an increase in the value of the awards. This 2018 report eloquently displayed the dissatisfaction among graduate students in regard to the value and number of federal scholarships, as well as the perceived financial security that comes with obtaining these scholarships.

During the COVID-19 pandemic, financial concerns of graduate students became even more prominent, with Statistics Canada reporting that over two-thirds of continuing postsecondary students were very or extremely concerned about the financial impacts of the pandemic (Wall 2020). The Toronto Science Policy Network's *COVID-19 Graduate Student Report* has also revealed that finances were at the forefront of worries for graduate students during the pandemic (Toronto Science Policy Network 2020).

These financial concerns for graduate students have continued, with inflation recently hitting a 40-year high, causing an increase in everyday expenses (Scherer and Smith 2022). While inflation affects all those in Canada, graduate students can be considered a particularly vulnerable group because their stipend values do not reflect average living costs. The Canadian housing market has also witnessed a large increase in real-estate prices. While the majority of graduate students are not in a position to purchase their own homes, rental prices across Canada are increasing at unprecedented rates (Chen 2022).

Given these pressures, graduate students have been actively advocating for increases to graduate student scholarships and funding in Canada. Most notably, our fellow student-led science policy networks have been advocating for increased financial support for graduate students for many years. In response to the release of the *Fundamental Science Review in 2017,* commonly known as the Naylor Report (Advisory Panel for the Review of Federal Support for Fundamental Science 2017), the SPE started a campaign called #Students4TheReport that urged the government to implement the report's 32 recommendations. The SPE then launched a survey entitled *Rethinking Federal Research Funding* that investigated the opinions of graduate students and postdoctoral fellows on federal awards from the Tri-Council Agencies (SSHRC, NSERC, and CIHR). This report became a key piece of evidence in the formation of the Canadian Council of Academies' *Degrees of Success* Report that investigated barriers for PhD students when transferring into the labour market (Council of Canadian Academies 2021).

Despite this work, there is a lack of national data on graduate student finances, especially for the vast majority of graduate students who do not hold a federal scholarship. This survey aimed to investigate the state of graduate student finances in Canada. From 1305 responses of graduate students representing various geographical locations, years of study, fields of education, and demographic backgrounds, we examined costs, tuition, debt, sources of funding, reported financial struggle, and other financial indicators. As a result of these analyses, we propose six recommendations to better support graduate students and enable them to more fully focus on research. Our findings will be essential in implementing changes to support the Canadian research ecosystem. In addition, by sharing and bringing awareness to the stories and financial struggles of graduate students, we hope to humanise this issue.

Materials and methods

Survey development

The "National Graduate Students Finance Survey" was conceptualised and designed by members of the Ottawa Science Policy Network (OSPN) in collaboration with additional authors. The goal of the survey was to generate a comprehensive report on graduate student finances to use this information as leverage following the 2021 Federal Election and to launch a campaign to improve the well-being of graduate students in Canada. We aimed to report on the current status of stipends for master's and doctoral students as funded through the Tri-Council Agencies: the CIHR, NSERC, and SSHRC. Survey questions were carefully designed and best efforts were made to frame the questions objectively. The survey questions can be found in the supplementary file. According to the Tri-Council Policy Statement (TCPS-2), ethics review/approval was deemed unnecessary as "ethics review/approval is not required for projects that are considered Quality assurance and quality improvement studies, as these don't typically meet the purpose of research." Article 2.5 of the TCPS-2 states "Quality assurance and quality improvement studies, program evaluation activities, and performance reviews, or testing within normal educational requirements when used exclusively for assessment, management or improvement purposes, do not constitute research for the purposes of this Policy, and do not fall within the scope of REB review."

Survey distribution

From October to December 2021, using a list of all Canadian universities, we collected emails from graduate student associations, societies, departmental groups, and clubs that were accessible online. The survey was sent out by email between 10 December 2021 and 14 December 2021, with follow-up emails sent in the following weeks. The survey was open from 10 December 2021 until 14 April 2022. To reach a broad audience, the survey was advertised on social media platforms, including Twitter and Instagram. In line with our organisation's commitment to bilingualism, our survey was accessible in both of Canada's official languages (English and French).

Data policy

Responses to the survey are held as confidential by the OSPN, a student organisation at the University of Ottawa. Aggregated, anonymous, or de-identified data may be shared with representatives of student associations, university administrations, advocacy groups, legislatures, and governments across Canada, including data subsets where appropriate.

Although we did not collect identifying information (e.g., respondent names), individuals may still be identifiable by

combining data. To protect anonymity, we will only publish or share subsets of data generated by combining responses of two or more questions when there are at least five responses in all categories. Any numerical data will only be published or shared as summary statistics (for example, mean and standard deviation) or as binned data (such as a histogram). Responses will never be shared or published in such a way that links together all an individual's responses.

Raw data collected in the survey will be retained for a period of seven years after which it will be deleted. Any data published or shared with other organisations will be retained indefinitely. Once the analysis of the data has concluded, access will only be granted to the OSPN president (or other OSPN members that the OSPN president determines are responsible for data custody).

Data preprocessing

Data preprocessing occurred prior to analysis. Based on the unique identifiers provided for each survey entry and demographic information, 14 respondents were determined to be duplicates (i.e., the same individual filled in the survey twice) and were removed. As data analysis was performed in English, all French responses were translated into English. Numerical data were transformed into a consistent format (e.g., removing comma separators and $ symbols), but no numerical values were changed. In cases where the authors felt there was an obvious entry error or possible misunderstanding of the question, the responses were excluded from the analyses.

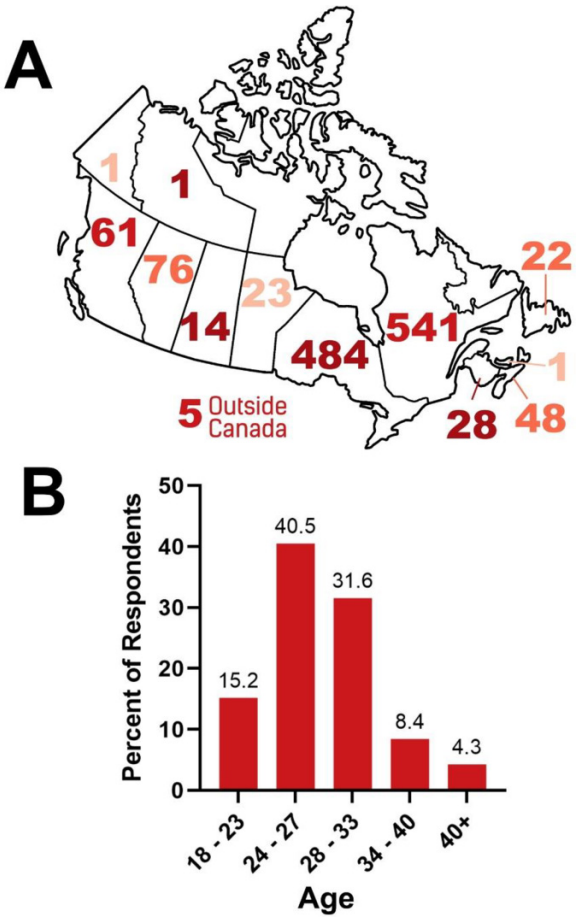
For some questions, where a free-form response was an option, responses were classified into groups to better enable comparisons. This was done for the following survey sections: current level of study Q2, ethnic identities Q8, gender identity Q9, field of study Q14, living environment (urban, suburban, rural) Q23, ownership status of accommodation Q24, who else the respondent lived with Q25, and sources of financial support Q27. In these classifications, effort was taken to retain the meaning of a response, especially for demographic information, and reclassification was based on all information available and best judgement. A copy of the unprocessed survey answers has been retained.

Statistics and data visualization

Throughout this report we used a significance threshold of 5% (i.e., $P < 0.05$). This means that for findings to be claimed significant, there is a greater than 95% probability that any difference detected is not solely due to random variation in sampling. However, it does not provide any information about why two (or more) groups may differ and does not take into account possible underlying factors—only that the difference is unlikely to be due to statistical chance alone. For comparing categorical data, the χ^2 test was performed using the categories presented in the plot or other summary data. For numerical data comparisons, a two-sample Kolmogorov–Smirnov (KS) test or Kruskal–Wallis (KW) test was used to compare the means when more than two categories existed. When comparing more than two categories, the statistical test does not provide information about how or which categories differ—only that the total extent of differences seen are unlikely to be due to randomness of sampling.

For data visualisation, bar graphs were created using GraphPad Prism 9.5.0; donut charts using R 4.2.1 (R Core Team 2022) and the lessR package (Gerbing 2021); and violin and QQ-plots were created using Python 3.8.10 with the statsmodels package (Seabold and Perktold 2010).

Fig. 1. Geographic location and age: (A) number of respondents from each province and territory ($n = 1305$) and (B) age category of respondents by % ($n = 1305$).



Results and discussion

Demographics of survey respondents

One aim of our survey was to assess the financial situation of graduate students across Canada, including students from varying programs, fields of study, and ages. We received a total of 1305 responses, 1030 of which were in English and 275 in French. Nearly all respondents were registered as full-time students (96.7%), with only a small percentage registered in part-time programs (2.3%). Students that resided in QC and ON made up 78.6% of respondents, with 484 respondents from ON and 541 from QC (Fig. 1A). This is consistent with Statistics Canada data from the National Graduates Survey that showed QC and ON made up 73.8% and 73.7% of master's and PhD graduates, respectively (Statistics Canada 2019). Ad-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

ditionally, over 70% of respondents were between the ages of 24 and 33 (Fig. 1B). Again, this is consistent with the Statistics Canada data that showed an average graduate student's age was 31 and 35 upon graduation from master's and PhD programs, respectively (Statistics Canada 2019). This provides confidence that the conclusions made in this survey are a valid representation of the overall graduate student profile in Canada.

A majority of respondents (57.6%) were pursuing a doctoral degree and nearly all were in a research-stream program (Fig. 2A). Similarly, the majority (60.9%) of master's student respondents were in the research stream, while the remaining were in course-based (19.8%) and professional master's programs (19.3%) (Fig. 2A). Responses across years of study were evenly dispersed (Fig. 2B). Survey respondents were primarily in science-based programs, including physical science (26.3%), medical sciences (23.8%), life sciences (23.5%), and social sciences (16.4%) (Fig. 3). Arts and humanities students comprised 7.7% of respondents (Fig. 3). A small percentage of business students (1.6%) and law students (0.5%) also responded to the survey (Fig. 3). As such, our overall findings are most representative of graduate students in STEM-related fields, which encompassed 73% of our respondents. This is higher than the 46.4% of students in science-related graduate degrees, as shown in the Canadian Association for Graduate Studies 2019 survey. This survey also estimated that there were about 63 000 graduate students in Canada. Additional demographic data can be found in Supplementary Table 1.

Tuition remains a large financial burden for graduate students

Tuition is one of the largest expenses for graduate students in Canada and is primarily regulated by provincial governments. According to Statistics Canada, the average tuition for Canadian graduate students is about $7437 per year (up 1.7% from last year) (Statistics Canada 2022*a*). NS graduate students have the highest tuition, on average $10 591 per year, with ON ($9385) and BC ($9994) close behind. Comparatively, QC has the lowest average tuition for graduate students, at an average of $3582 per year (Statistics Canada 2022*a*).

Universities can use different tuition fee structures, with various opportunities for tuition rebates and waivers. Tuition freezes, which were implemented in ON during the COVID-19 pandemic, can also affect the rising tuition costs across the country. In addition to tuition fees, students are required to pay ancillary fees throughout their education. These can include fees for fitness or athletic services, technology use, public transit, health and dental plans, etc.

To assess the financial situation of graduate students, we investigated tuition variance among students. We found that the average annual tuition paid by respondents across Canada was $7518 for domestic students (see the analysis of international student costs below). The lowest average cost of tuition was reported in NL at $5016.27 per year, while the highest average was reported in ON at $8972.46 per year, with NS following behind at $7770 (Fig. 4). The highest tuitions were among four provinces in both our survey and the Statistics Canada data: ON, NS, NB, and BC.

Tuition remains a large burden on graduate student finances, though it is questionable whether these costs are factored into university funding packages. For example, one respondent stated "My stipend total this year (including additional TA work) was $29 000. From that, ~$6000 was deducted for tuition. $23 000/year is not enough to live on—this is below the low income threshold that qualifies you for assistance program (ex: Leisure Access Pass) in Edmonton. Except as students we are exempt from these programs."

To further characterise tuition distribution among respondents, we asked graduate students to report their out-of-pocket payments. This accounts for students' tuition waivers, which are offered through universities to subsidise the cost of tuition. Across domestic students, the average out-of-pocket tuition was $3226. Our results reflect a bimodal distribution, with the vast majority of respondents either paying 0%–10% or 90%–100% (Fig. 5), where 82.6% of domestic graduate students either pay less than 10% or more than 90% of their tuition out of pocket, highlighting an all-or-none approach to tuition finances. This variation in out-of-pocket tuition payments could partly be attributed to money that some students receive through scholarships. When broken down by scholarship status, we found that domestic respondents who held a government award paid significantly less tuition out of pocket (KS test, $P < 10^{-5}$), with 70% of award holders paying 10% or less of their tuition out of pocket compared with 37.3% for those who did not hold a government award (Fig. 6). However, it is important to note that the bimodal distribution was still seen for students who do not hold a government award, suggesting that other factors also contribute.

We also hypothesised that this trend may impact the financial situation of graduate students paying either 0%–10% or 90%–100% of their tuition out of pocket. To assess this, respondents in either group were cross-compared with how they ranked their financial situation (Fig. 7). We found a significant difference (chi2 test, $P = 0.011$) between these two groups, with those who paid over 90% of tuition out of pocket more likely to be struggling financially.

Graduate students in Canada struggle to pay many costs of living

As described in the introduction, Canada is currently experiencing a 40-year inflation high, resulting in higher costs of living for all Canadians. In combination with the lingering effects of the COVID-19 pandemic, graduate students are particularly financially vulnerable, considering slow and inflexible funding models. To better understand how graduate students are spending their finances, we investigated their housing costs, area of residence, living situation, and monthly expenses.

Our findings show that over 80% of respondents were renting, with less than 10% owning homes (Fig. 8A). This is concerning compared to recent Statistics Canada data profiling home ownership. For young adults aged 25–29 years, 36.5% owned their homes. For those aged 30–34 years, homeownership rates were 52.3% (Statistics Canada 2022*b*). This indicates a large financial burden experienced by graduate students compared to their peers. One respondent remarked:

Fig. 2. Level and year of study: (A) program and current study level of survey respondents by % ($n = 1303$). The doctoral category includes professional doctoral (e.g., EdD, PharmD, and DBA), research stream doctoral (PhD), course-based doctoral, and MD PhD. (B) Current year of study of survey respondents by % ($n = 1297$).

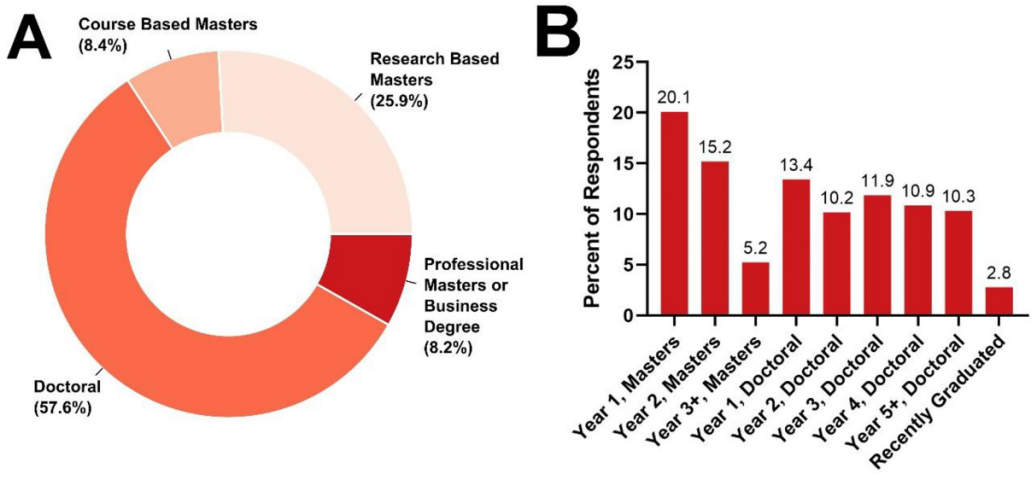


Fig. 3. Field of study. Respondents self-reported their field of study from seven categories: arts and humanities, business, law, life sciences, medical sciences, physical sciences (engineering, physics, mathematics, etc.) and social sciences ($n = 1305$). If respondents indicated multiple fields of study or an option not listed in the survey, they were categorized into one of the seven categories based on their answers for analysis purposes.

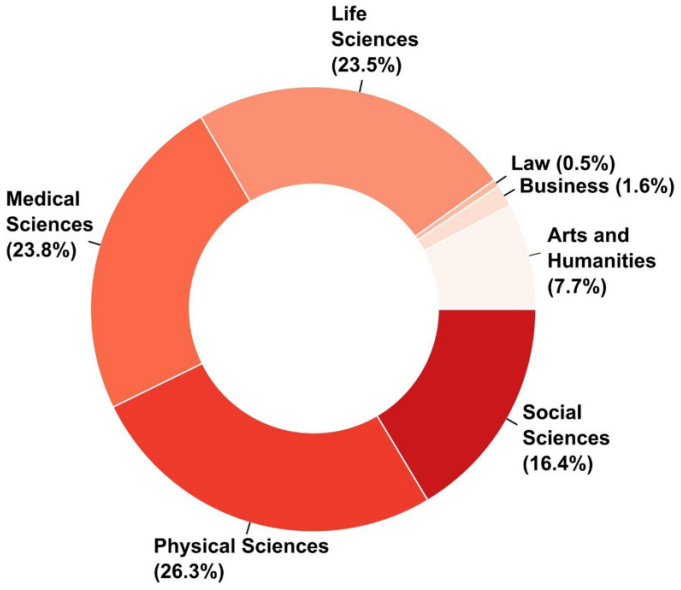
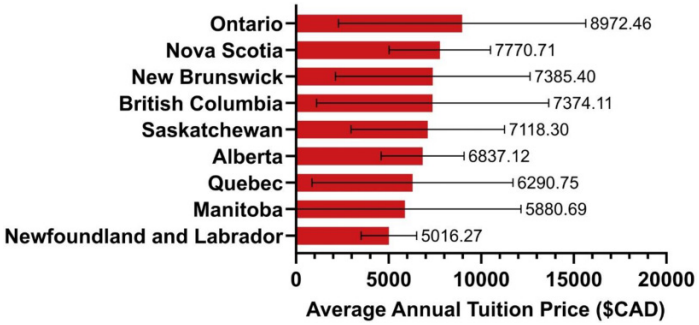


Fig. 4. Annual tuition fees. Average annual tuition reported by respondents in Canadian dollars by province ($n = 886$). Categories that had less than five responses were excluded (PE, NT, YT, and outside of Canada). Bars represent mean values ± SD.



"I had to take cash advances on my credit card and borrow money from family to get by… As such I am still paying off debts which has delayed accomplishing milestones (like buying a house and starting a family) that others my age are currently doing. Bottom line graduate students should be paid at least a minimum wage full-time salary as taking on secondary jobs to supplement income is just not possible if we want to graduate on time" and another respondent noted that "after groceries and rent, [their] funding leaves [them] with a negative cashflow position." While others age-matched are building wealth and investing in their future, graduate students are subjected to the insecurity of the rental market.

Three-quarters of respondents lived in urban areas (74.5%), while the remainder lived in suburban areas (20.4%) and in rural areas (4.7%) (Fig. 8B). Most respondents lived with other people, such as a partner, roommate(s), or family (Fig. 8C), while 25% lived alone. A large majority of our respondents did not have dependents (Fig. 8D), while the remaining 13% had one or more dependents (further analysis of those with dependents can be found further below).

Based on those who paid rent, a quarter of respondents (25.3%) paid $500–$750 monthly (Fig. 8E). Most respondents paid between $500–$1500 in rent, a wide range that depends on location, roommate situation, and other factors. One percent of respondents (14/1305) did not have stable housing.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 5. Domestic respondent percent tuition paid out of pocket. Percent of tuition paid out of pocket after any applicable scholarships for domestic respondents ($n = 885$). Tuition paid out of pocket was divided by the total annual tuition fees for each respondent to calculate the %.

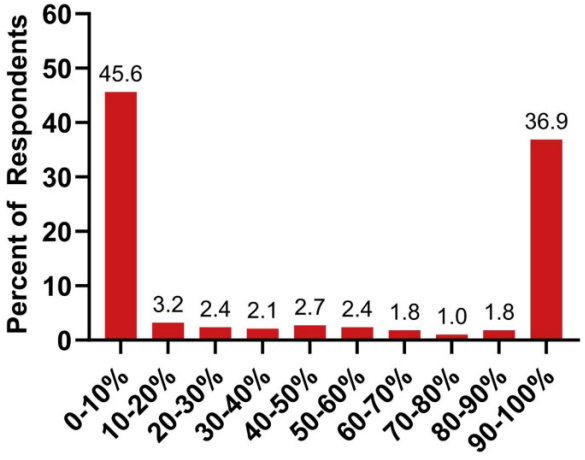
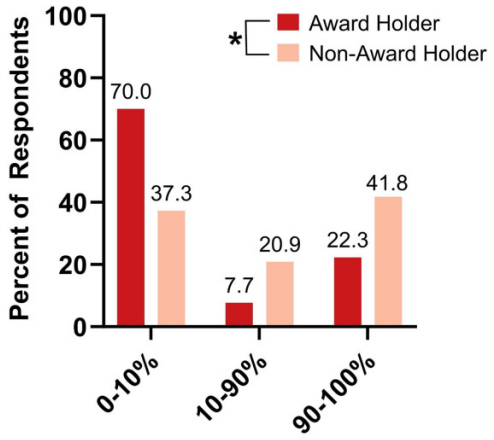


Fig. 6. Domestic student percent tuition paid out of pocket by award holder status. Percent of tuition paid out of pocket after any applicable scholarships for domestic students with a government award ($n = 220$) and without a government award ($n = 665$). Tuition paid out of pocket was divided by the total annual tuition fees for each respondent to calculate the percentage. $^{*}P < 0.05$, KS test (award holders vs. non-award holders).

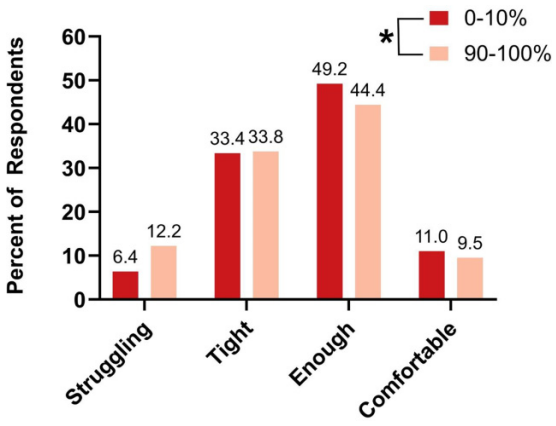


Monthly expenses for respondents were surveyed across different categories (Fig. 9). The highest expense, aside from rent (Fig. 8E), were groceries, with $100–$400 spent monthly by over 70% of respondents.

Canadian graduate students are experiencing financial stress

Between low pay, increasing costs of living, and accumulating debt, it is clear that financial stress is a major concern for many graduate students. Not only does financial stress impact graduate students' well-being and ability to complete their studies, but insufficient savings and outstanding debt can impact their future financial security. This has become a dominant narrative across research ecosystems as many students are speaking out on these financial struggles in Canada (Marhnouj 2022) and in other countries (Woolston 2022).

Fig. 7. Financial situation of low (0%–10%) vs. high (90%–100%) % tuition paid out of pocket respondents ($n = 1004$). $^{*}P < 0.05$, χ^2 test (0%–10% vs. 90%–100%).

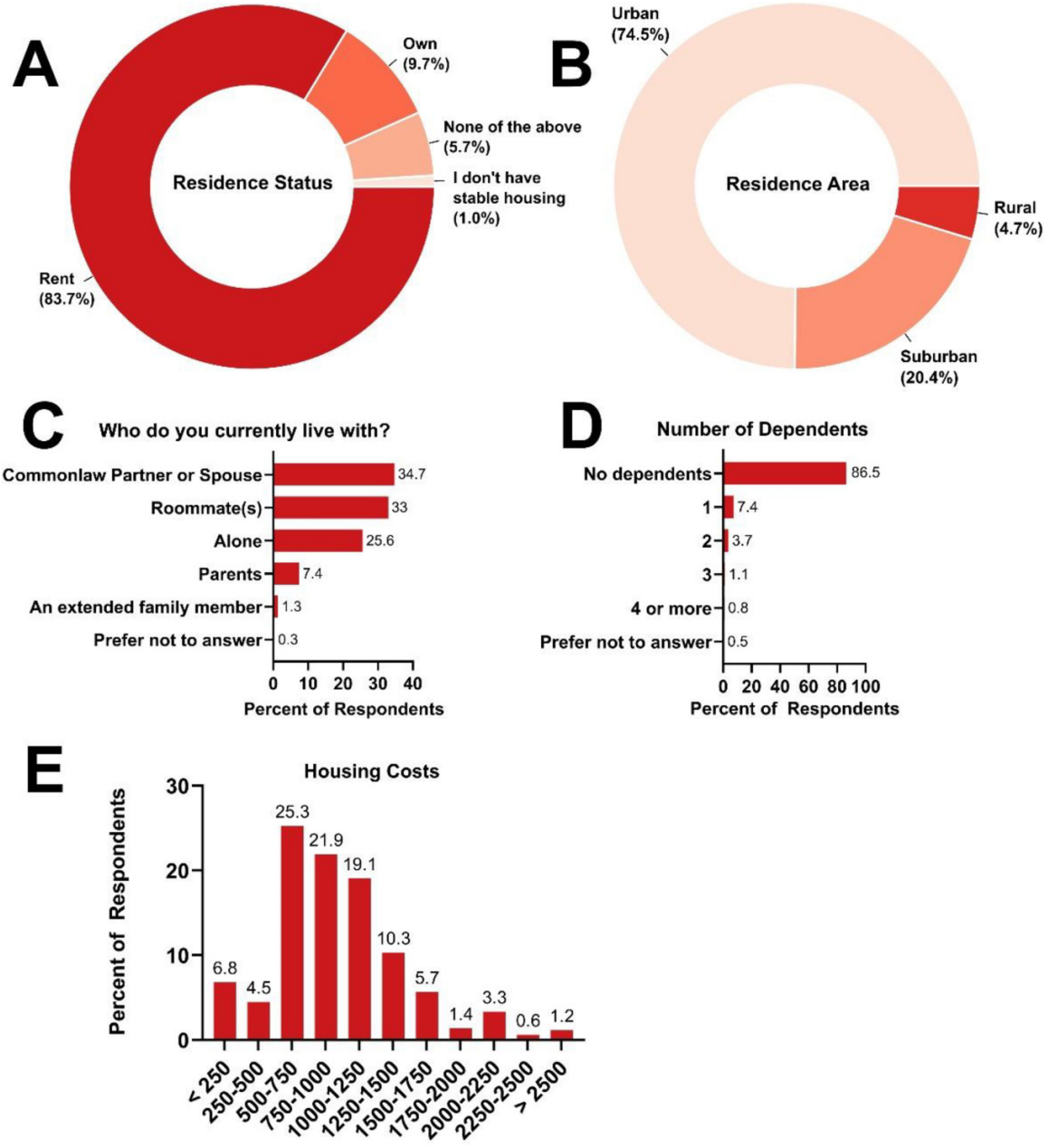


To investigate this question in graduate students across Canada, we asked survey respondents to identify their comfort levels with their current financial situation. Our team examined how concerned students were about various financial stresses, including living expenses, debt, and emergency funds. To investigate students' financial status, we asked participants to choose which most appropriately defined their current financial situation: comfortable (I am very comfortable financially. I have enough money that I do not need to worry about monthly expenses and spend as I would like.), enough (I have enough. My finances are a bit tight but I live within my means and can afford to provide for myself.), tight (every month is tight, but I am getting by. I am often making sacrifices to pay for necessities.), and struggling (I am struggling financially. I often do not have enough to make ends meet.) Nearly half of all respondents (43.4%) identified that they were either frequently struggling to make ends meet or that their tight finances forced them to make sacrifices to afford necessities (Fig. 10A). Less than 10% of students stated that they were very comfortable with their current financial situation (Fig. 10A). Unsurprisingly, those who identified as struggling to make ends meet generally had the lowest stipend values, whereas those who identified as being very comfortable with their financial situation had the highest (Fig. 10B). One student even stated that they had to "drop out of [their] Masters program due to finances and having to work full time to afford the cost of living." This student also highlighted how the "faculty [they were] in did not support students working externally yet offered very few alternatives to afford both tuition, living expenses, and debt payments."

The majority of respondents (85.7%) expressed experiencing stress/anxiety about finances at some point during their graduate studies. Meanwhile, 54.7% experienced difficulty

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 8. Residence situation: (A) residence status of respondents by % ($n = 1305$); (B) area of residence by % ($n = 1305$); respondents who listed areas other than urban, rural, or suburban were re-categorized into one of the three; (C) current living arrangements of respondents (multiple selections possible) by % ($n = 1305$); (D) number of dependents reported by respondents by % ($n = 1305$); and (E) monthly housing costs ($CAD) by % of respondents ($n = 1291$).



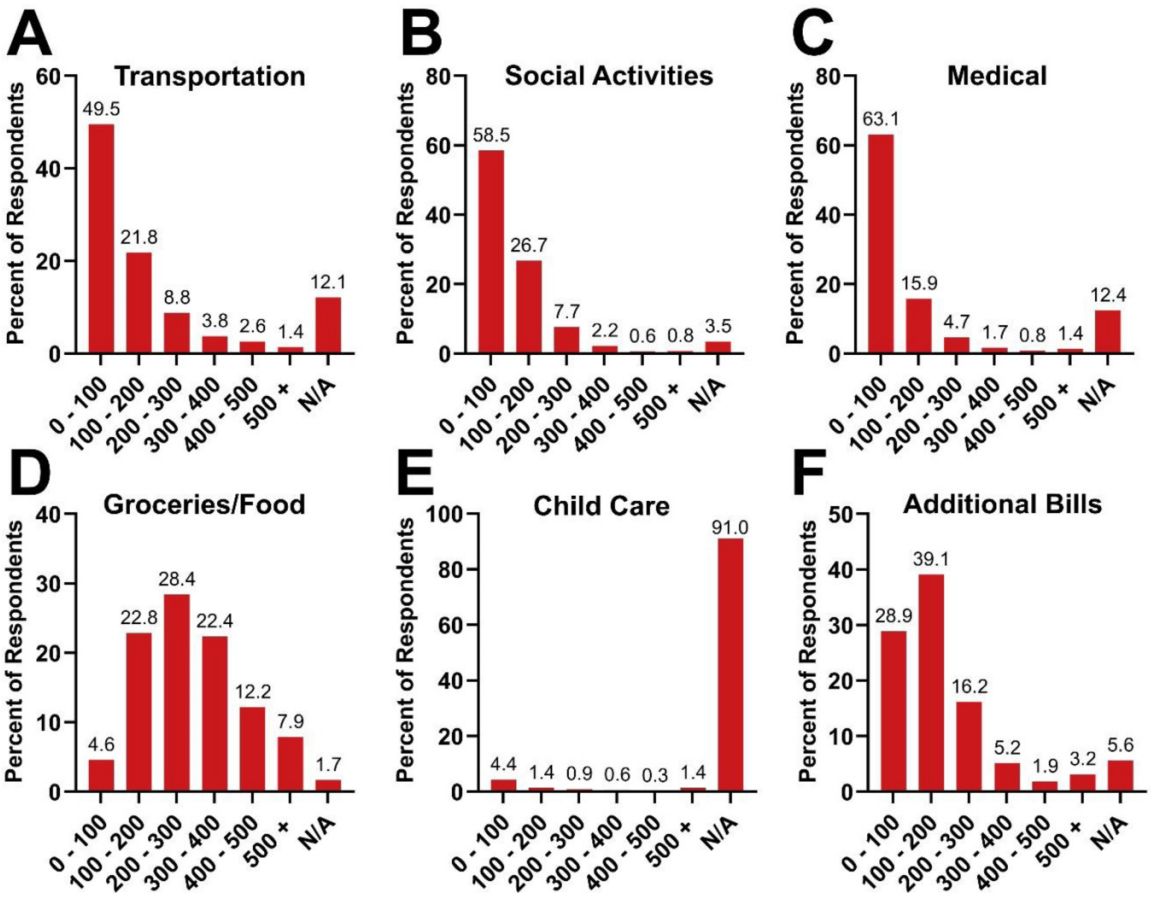
budgeting, and 30.7% have considered leaving their studies due to financial struggles alone (Fig. 11A). Concern about insufficient funds for living expenses was common among respondents, with over 30% having experienced concern about rent and food (Fig. 11A) and 27.2% either "always" or "often" worrying about their ability to pay for bills (Fig. 11B).

The standard advice from financial experts and the recommendation by the Financial and Consumer Agency of Canada is to ideally have at least 3–6 months of savings for living expenses (Financial Consumer Agency of Canada 2022). Only half of our survey respondents met this recommendation, with 52.3% having only 0–3 months of living expenses saved (Fig. 11C) and 53.7% of students either "always" or "often" worrying about their ability to pay for emergency expenses (Fig. 11B). These findings highlight that graduate students are vulnerable to unexpected financial challenges.

As an expected consequence, graduate students worry about their future financial security. Nearly 60% of respondents indicated that they either "always" or "often" worry about future financial security, and a similar number (57%)

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 9. Monthly Expenses. Respondents were asked to share their average monthly expenses based on range values in each category: (A) transportation ($n = 1305$), (B) social activities ($n = 1305$), (C) medical ($n = 1305$), (D) groceries/food ($n = 1305$), (E) child care ($n = 1305$), and (F) additional bills ($n = 1297$). Data are shown by the % of respondents for each category.



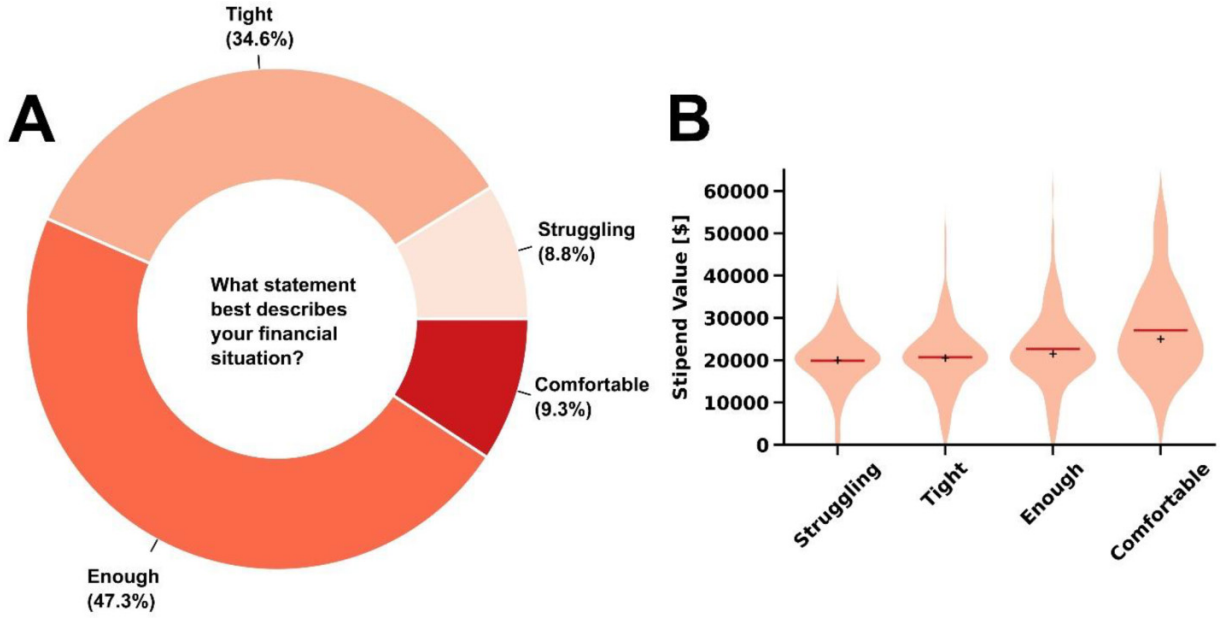
indicated "always" or "often" worrying about their ability to obtain a job in the future (Fig. 11B). A key contributor to this may be the accumulation of debt throughout one's university studies, as over half of respondents (52.2%) had some outstanding debt, the majority of which had over $10 000 owed (Fig. 12A). Indeed, respondents with outstanding debt reported worrying about future financial security more often than respondents with no outstanding debt ($P < 10^{-5}$, χ^2 test) (Fig. 12B). A higher percentage of respondents with outstanding debt also reported struggling financially compared to those with no outstanding debt ($P < 10^{-5}$, χ^2 test) (Fig. 12C). Reliance on debt seems to partly stem from a lack of parental support, as only 20% of students with debt received financial support from parents, compared to 30% of students with no debt having received parental support ($P = 0.000028$, χ^2 test) (Fig. 12D). Additionally, doctoral students had a higher amount of debt on average than master's students ($34 370 for doctoral vs. $29 272, $24 942, and $28 903 for course-based, research-based, and professional and business master's students, respectively; Fig. 13A). This is unsurprising given the increased length of study, but note that master's students and doctoral students still reported similar responses regarding their financial situations (Fig. 13B).

Graduate students in Canada often have to work outside of their studies

Non-academic work completed outside of a graduate student's studies, to supplement income, is often considered taboo as it can be frowned upon or, in some cases, is against university regulations. Academic institutions can prevent graduate students from accessing alternative forms of income by restricting the number of hours they are permitted to work outside of their studies. For example, McGill University allows graduate students up to a maximum of 12 h work per week outside of their studies to maintain full-time status (McGill University 2022). In 1994, the Ontario Council on Graduate Studies, of which 21 universities across ON are members, implemented the 10 h rule, which restricts the number of hours graduate students can work outside of their studies to 10 h per week (Ontario Council on Graduate Studies 2021). Scholarships issued by the Ontario Government also adhere to this rule and regulate the amount of time another establishment can employ a student. There is growing awareness of the problems arising from the 10 h rule. For example, the University of Waterloo recently cited that there was a "strong desire to provide greater flexibility in employment hours" and changed the requirements of the 10 h rule to now allow for a student to work up to 20 h per

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 10. Financial situation. (A) Respondents were asked what best describes their financial situation: comfortable (I am very comfortable financially. I have enough money that I do not need to worry about monthly expenses and spend as I would like.), enough (I have enough. My finances are a bit tight but I live within my means and can afford to provide for myself.), tight (every month is tight, but I am getting by. I am often making sacrifices to pay for necessities.), and struggling (I am struggling financially. I often do not have enough to make ends meet.). Data are shown by % ($n = 1303$). (B) Violin plot of respondents' financial situation and stipend value ($n = 1303$). The red line and + sign represent the mean and median stipend value of respondents in each group, respectively.



week (University of Waterloo 2022). In our survey, one respondent noted that "[m]ost graduate students in our department work full-time jobs on top of completing their studies AND TA [teaching assistant] or RA [research assistant] positions. The department complains about students not finishing their programs within a short amount of time, yet they won't pay us more and strongly advise against/look down on students who work over the 10 hour/week 'limit'." Abolishing the 10 h rule has increasingly become a topic of conversation concerning financial struggle among graduate students. In provinces outside of ON, this is less of an issue as many universities/provinces have higher limits on work outside of studies (for example, the University of Winnipeg has a limit of 35 h per week).

To investigate external employment of graduate students and gain insight into their work–study balance, we asked respondents whether they were externally employed (or get paid for a service) outside of their graduate studies. This included any money that is made in addition to their stipend. Many graduate students work as TAs, and this was counted as additional work if the payment was not included or required as part of their stipend.

More than half of all respondents (55.6%) worked in addition to their studies (Figs. 14A, 14B). Of this majority, nearly half worked less than 10 h per week, while the other half worked over 10 h per week. This includes 18.5% who worked between 5 and 10 h/week and 17.8% who worked between 10 and 20 h/week (Fig. 14C). A small % of respondents (6.5%) worked over 30 h per week.

Additionally, when looking at financial struggles, respondents who worked were significantly more worried about their ability to pay back student debt compared to those who did not work (χ^2 test, $P = 0.036$) (Fig. 15). Other financial concerns were similar between the working and non-working groups.

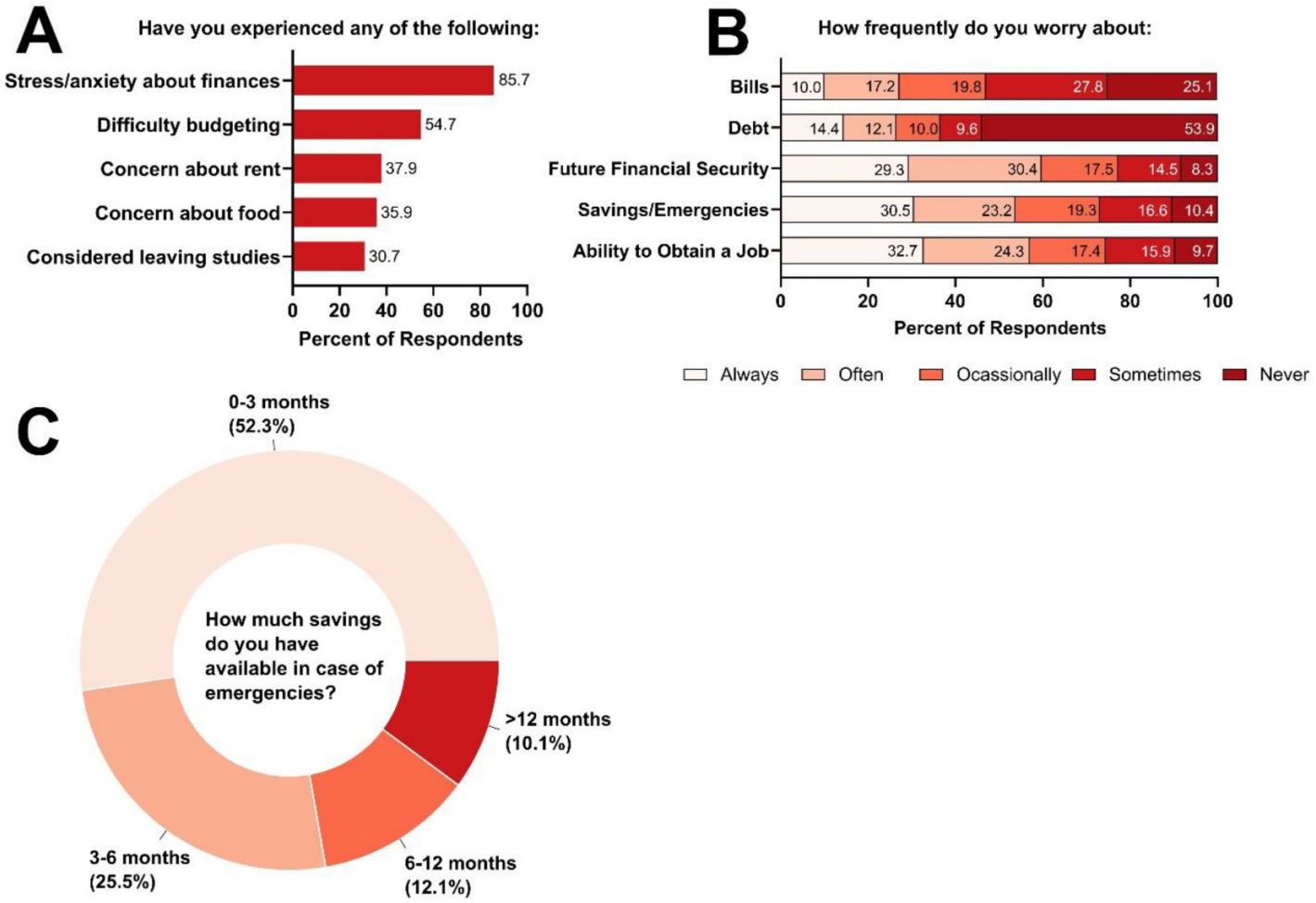
Graduate students experience errors in stipend payments

Graduate students depend on their stipends for living costs; however, some graduate students have experienced "stipend errors," defined as a disruption in the regularly scheduled payment of a stipend. Our survey sought to understand the stipend errors experienced by graduate students using questions based on a study conducted by the Dalhousie Working for Inclusion in Chemical Sciences Group (Dalhousie Working for Inclusion in Chemical Sciences Group 2021).

When asked whether survey respondents experienced any errors regarding stipend payments, over 60% of respondents indicated they had not experienced any errors (Fig. 16A). In contrast, 27% of respondents stated they received late stipend payments. Students also had errors in getting paid too little, too much, or had their stipend payments put towards a uni-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 11. Financial stressors. (A) Concerns and stress experienced by respondents (multiple selections possible) by % ($n = 1178$). (B) Respondents were asked how frequently they worry about their ability to obtain a future job in their chosen field, savings/ability to pay for emergency expenses, future financial security, ability to pay back their student debt, and ability to pay bills. Data are shown by % in each category ($n = 1305$). (C) Amount of savings, measured by months of living expenses, that respondents have available for emergencies, by % ($n = 1264$).



versity expense/fee (Fig. 16A). More than one in three respondents (36.4%) have experienced at least one error in stipend payments (Fig. 16B).

Personalized responses demonstrated that stipend errors can seriously affect graduate students' finances, time, and energy. For example, one student noted they were "not enrolled by the financial department to receive payment for over two months," while other students had stipend payments that were anywhere from 8 weeks to 3 months late. One student said their stipend was "mistakenly cancelled" and it took "a lot of back and forth with profs and admin" to receive it. Students also stated that their "stipend payment was put towards a university expense/fee without [their] prior knowledge."

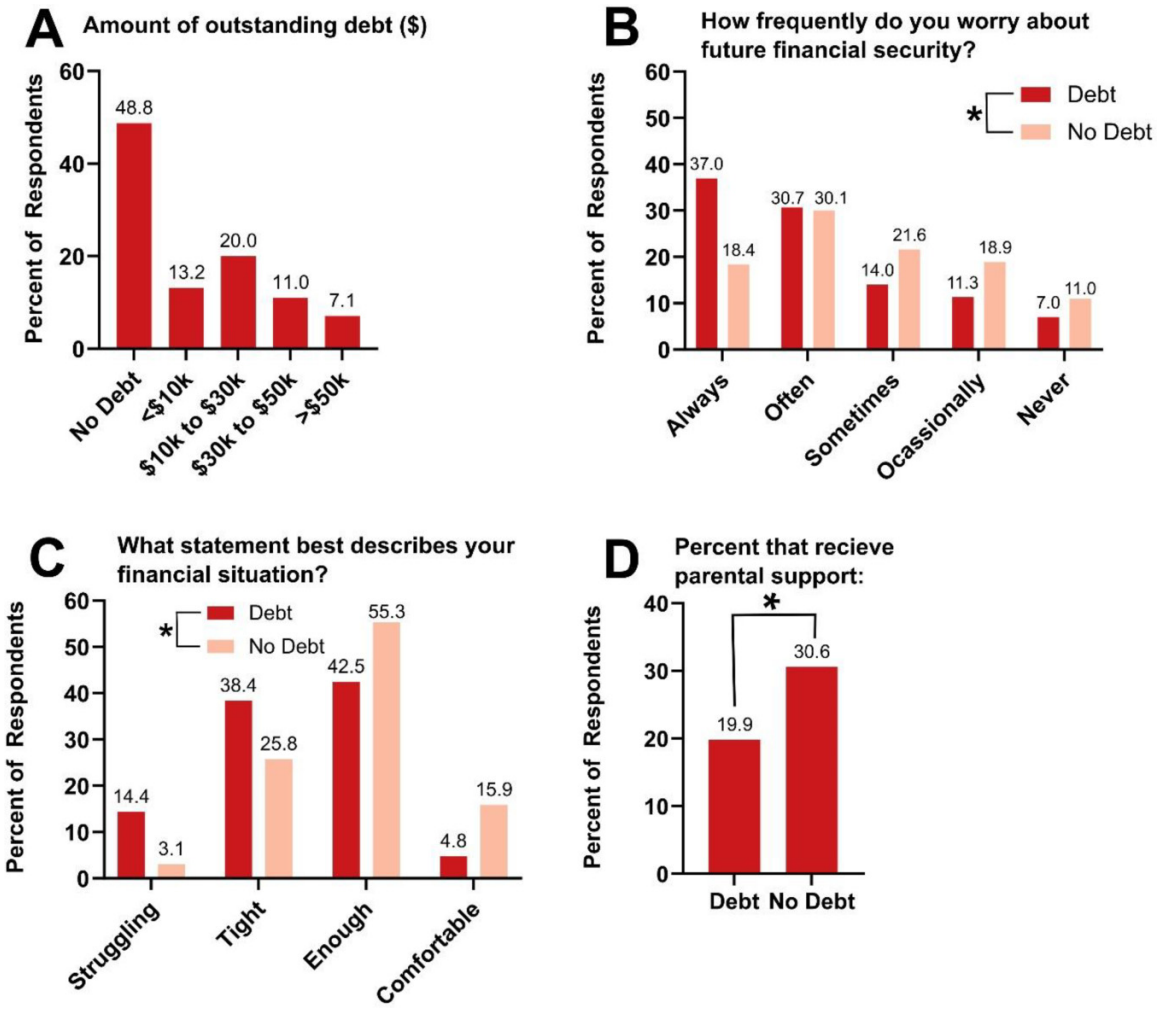
Since this process largely depends on the university administration that is responsible for stipend payments, these analyses will need to be assessed on a university and(or) department level basis to properly determine the impacts of stipend errors. While our results show that stipend errors occur, future work should identify why this is happening and how this can be improved to alleviate student financial stress in this area.

Graduate students are not always reimbursed in a timely manner for out-of-pocket research costs

While the cost of living continues to increase at record rates and stipends remain stagnant, many graduate students are not equipped to deal with unexpected expenses. With only 9.3% of respondents having felt comfortable financially (Fig. 10A), we were concerned about the impact of upfront research expenses paid by graduate students for their financial security. These types of expenses can include purchasing lab equipment and paying for trips related to field work, software, parking, and academic conferences. Field work and conferences, in particular, can incur hundreds of dollars in expenses for travel fees. Additionally, these costs may require

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 12. Financial situation by debt status. (A) amount of outstanding debt of respondents in ranges by % ($n = 1130$). (B) Financial situation of respondents with debt ($n = 584$) or no debt ($n = 555$). (C) Financial situation of respondents with debt ($n = 584$) or no debt ($n = 555$). Data are shown by the % of respective debt status. (D) Parental support of respondents with debt ($n = 584$) or no debt ($n = 555$). $^{*}P < 0.05$, χ^2 test (debt vs. no debt).



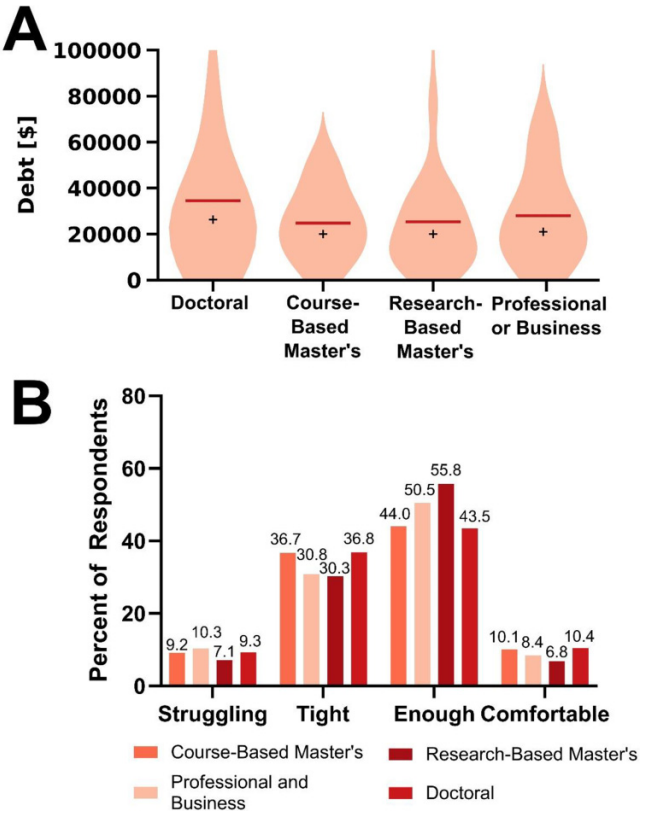
payment months in advance of a specific date. A delay in reimbursement, or absence of one, could lead to increased interest on credit card charges and(or) lack of money to pay for standard living expenses.

To determine the extent to which this situation may occur, we surveyed respondents on out-of-pocket academic expenses and the timing of their reimbursement. Over 50% of respondents, in doctoral (73.6%) or research-based master's degrees (52.5%), have previously paid for academic expenses upfront (Fig. 17A). For graduate students in course-based professional and business master's degrees, the percentage of respondents was lower at 34.81%. Within these proportions, we focused on expenses that needed to be paid out of pocket. Approximately 10%–15% of graduate students in a research-based master's (10.6%) or doctoral degree (14.6%) were required to pay these expenses. A higher proportion of students in course-based professional and business master's degrees (26.1%) were required to do the same. On average, students in course-based programs paid less in upfront academic expenses but were required to do so largely out of pocket or without reimbursement. As there may be some overlap in respondents who answered that they paid out-of-pocket expenses and those who answered that they failed to be reimbursed, further examination may be required for these concerns.

The vast majority of students waited longer than a week for reimbursements (Fig. 17B). Only 35.9% of students in research-based master's degrees, 39.9% of students in doctoral degrees, and 27% of students in course-based professional or business master's degrees were reimbursed within 4 weeks. Between 4% and 9% of respondents were required to wait more than 3 months to be reimbursed. Having to wait periods of over a month can impact financial security, as students have regular monthly expenses such as rent, utilities, credit card bills, etc. One student commented that "for conferences and plane tickets, [they] could only request a reimbursement after the conference has actually happened. This means I pay the conference fee and 6 months later I can get it reimbursed." Thus, waiting for reimbursements is likely

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 13. Financial situation by level of study. (A) Violin plot of average debt of research-based master's ($n = 337$) and doctoral respondents ($n = 750$) in Canadian dollars (CAD). The red line and + sign represent the mean and median stipend values of respondents in each group, respectively. (B) Financial situation of research-based master's ($n = 337$), course-based master's ($n = 109$), professional and business master's ($n = 107$), and doctoral ($n = 750$) respondents. Data are shown by the % of respective level of study.



to have impacted the ability of the survey's respondents to pay for their living expenses. Of these, 23.1% of students in course-based master's degrees, 13.4% of students in doctoral degrees, and 9.4% of students in course-based professional and business master's degrees have indicated this was the case (Fig. 17C).

This survey provides a cursory overview of the effects of upfront academic expenses on graduate students and has highlighted this as a real issue. Future work should aim to provide a more detailed picture of this issue and how it impacts students. For example, while we did not ask about the nature of these expenses, it may be valuable to delve further into the expenses students were asked to pay for, such as whether the expense was considered mandatory or not? Did students feel like they had a choice in the decision? What is the total amount of expenses students have paid for? Additionally, one of the risks with waiting for expenses is that if a student paid for it through a credit card and doesn't receive their reimbursement on time, then interest will begin to increase. This interest will not be covered by the eventual reimbursement.

Analysis of graduate student funding and support

Funding for graduate students in Canada is complex as it comes from various sources and is regulated on multiple levels. While some universities set minimum funding levels for student stipends, this can vary greatly between departments and even faculty within a university.

Over 70% of respondents declared that they received a stipend (Fig. 18A). The remaining 30% reported that they did not receive a stipend at all. Of these, 73% felt that they should receive a stipend (Fig. S3).

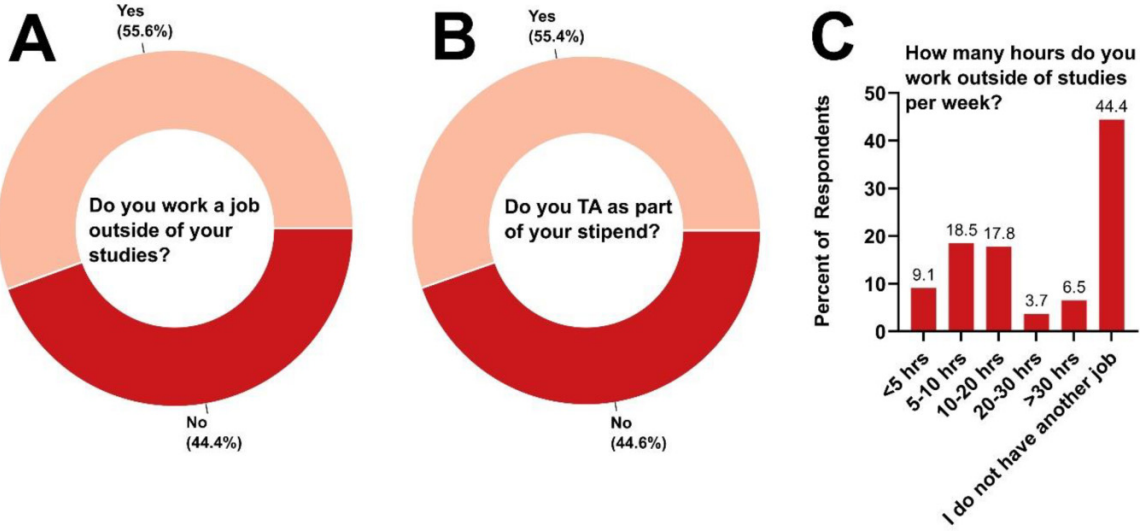
The average stipend reported was $23 750 for doctoral students, $19 725 for research-based master's students, $16 813 for course-based master's students, and $13 113 for business and professional master's students (Fig. 18B). If we consider the context of a typical 40 h work week, this becomes an average hourly pay of $9.18/h for a master's student and $11.43/h for a PhD student. This is well below the $15/h minimum wage implemented by the federal government (Employment and Social Development Canada 2022*a*) and does not account for the many graduate students who work over the average 40 h standard.

Based on the Likert scale of satisfaction (1 as very unhappy and 5 as very happy) (Likert 1932), almost half the respondents (46%) felt unhappy about their stipend value, which highlights an issue with current stipend values (Fig. 19). When asked about their ideal stipend values, respondents reported an average of $33 619 for doctoral students, $29 213 for research-based master's students, $26 912 for course-based master's students, and $22 705 for business and professional master's students (Fig. 20). The SPE's *Rethinking Federal Research Funding Report* found that respondents' ideal value of federal scholarships was $21 000 for a master's student and $35 000 for a PhD student (Science and Policy Exchange 2020). In our results, master's students reported a much higher ideal stipend value compared to their current average funding, which puts them well below the poverty line (Statistics Canada 2022*c*). Further, the ideal value of $21 000 from the SPE survey would still fall below the poverty line in the vast majority of the country.

We asked respondents to share the source of their stipend. The majority of respondents (67%) were funded through their supervisor (Fig. 21A). Less than half (45.2%) were funded from their department/faculty. While national and provincial awards exist, only 20% were funded through federal awards, and 11.5% of respondents were funded through provincial awards. Most respondents (77%) reported only having 1–2 funding sources (Fig. 21D). The length of funding also varied from under 6 months to 4 years (Fig. 21C). About 45.1% of students have experienced either an increase or decrease in their stipend value: 25.9% have received an increase, 6.5% have received a decrease, and 12.7% have received both an increase and a decrease at some point in their studies (Fig. 22A). The top-reported cause of a stipend increase was that the respondent received a scholarship (further information about scholarships can be found below). Otherwise, respondents reported that their supervisor gave a stipend raise or their student association ne-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 14. Work outside of studies: (A) respondents were asked whether they work a job outside their studies, by % ($n = 1305$), (B) respondents indicated whether being a Teaching Assistant (TA) was a condition of their stipend ($n = 914$), and (C) number of hours respondents worked outside of studies per week, by % ($n = 1305$).



gotiated to increase student stipends (Fig. 22B). Conversely, the top reasons for students' stipends decreasing were a change in TA hours, their scholarship ending, or their supervisors decreasing their funding (Fig. 22B). In general, scholarships, TA hours, and the supervisor's financial status were the largest determinants of student stipend values.

Inadequate finances affect graduate students in all fields of study

Graduate student experiences, especially regarding financial situations, vary greatly depending on the field of study due to differences in tuition prices, stipend values, work expectations, and scholarship access.

Statistics Canada reported in 2022/2023 that students' tuition varied substantially across degrees. For example, those studying business, management, and public administration paid the highest tuition, averaging $14 272 per year. Graduate students in STEM fields, such as computer and information sciences and engineering, were on the higher end of tuition costs, paying an average of $8902 and $7454 per year, respectively. Comparatively, humanities graduate students had lower tuition on average at $4753 per year (Statistics Canada 2022*b*).

We were interested in exploring how tuition, stipend payments, and work varied among graduate students in different fields of study and whether these contributed to differences in financial struggle. We asked students to report their annual tuition fees, including ancillary fees. Unsurprisingly, business students had the highest annual tuition (Fig. 23), which was similar to what was reported by Statistics Canada. Physical sciences, medical sciences, and law students reported annual tuition costs above $9000, while social sciences, life sciences, and arts and humanities students paid over $8000 per year (Fig. 23).

Whether students receive a stipend depends on the program and degree type. The majority of students in the life sciences, physical sciences, and medical sciences reported receiving a stipend: 87.6%, 77.6%, and 71.4%, respectively (Fig. 24A). In comparison, only 49.5% and 44.4% of students in the arts and humanities and social sciences, respectively, receive a stipend. Business students were least likely to receive a stipend, with only 33.3% reporting receiving one. This disparity between programs is likely a consequence of stipends typically only being granted to students in research-based programs, which tend to be more common in the natural sciences. While the average stipend value was consistent across programs (Fig. 24B), the most common sources for stipends differed. Funding by supervisors was the most common stipend source for life sciences, physical sciences, and medical sciences students, followed by funding from the department/faculty and then TAs (life sciences and physical sciences) or federal awards (medical sciences) (Table 1). In contrast, for arts and humanities and social sciences students, the department/faculty was the most common stipend source, followed by TAs and then federal awards (arts and humanities) and supervisors (social sciences).

As expected, a larger percentage of students in programs receiving fewer stipends, particularly arts and humanities and social sciences, reported working outside of their studies (71.3% and 77.1%, respectively) (Fig. 25). In comparison, students in programs that more frequently receive stipends (e.g., life sciences, medical sciences, and physical sciences) were less likely to report working outside of their studies (44.6%, 54.3%, and 47.7%, respectively).

Financial awards support graduate student research

As previously outlined, graduate students fund their studies through different sources, including fed-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 15. Work outside of studies—financial situation. How frequently do respondents who work or don't work worry about their (A) ability to pay bills, (B) ability to pay back their student debt, (C) future financial security, (D) savings/ability to pay for emergency expenses, and (E) ability to obtain a future job in their chosen field. Data are shown by % of respondents in each category: work ($n = 727$) and no work ($n = 578$). $^{*}P < 0.05$, χ^2 test (work vs. no work).

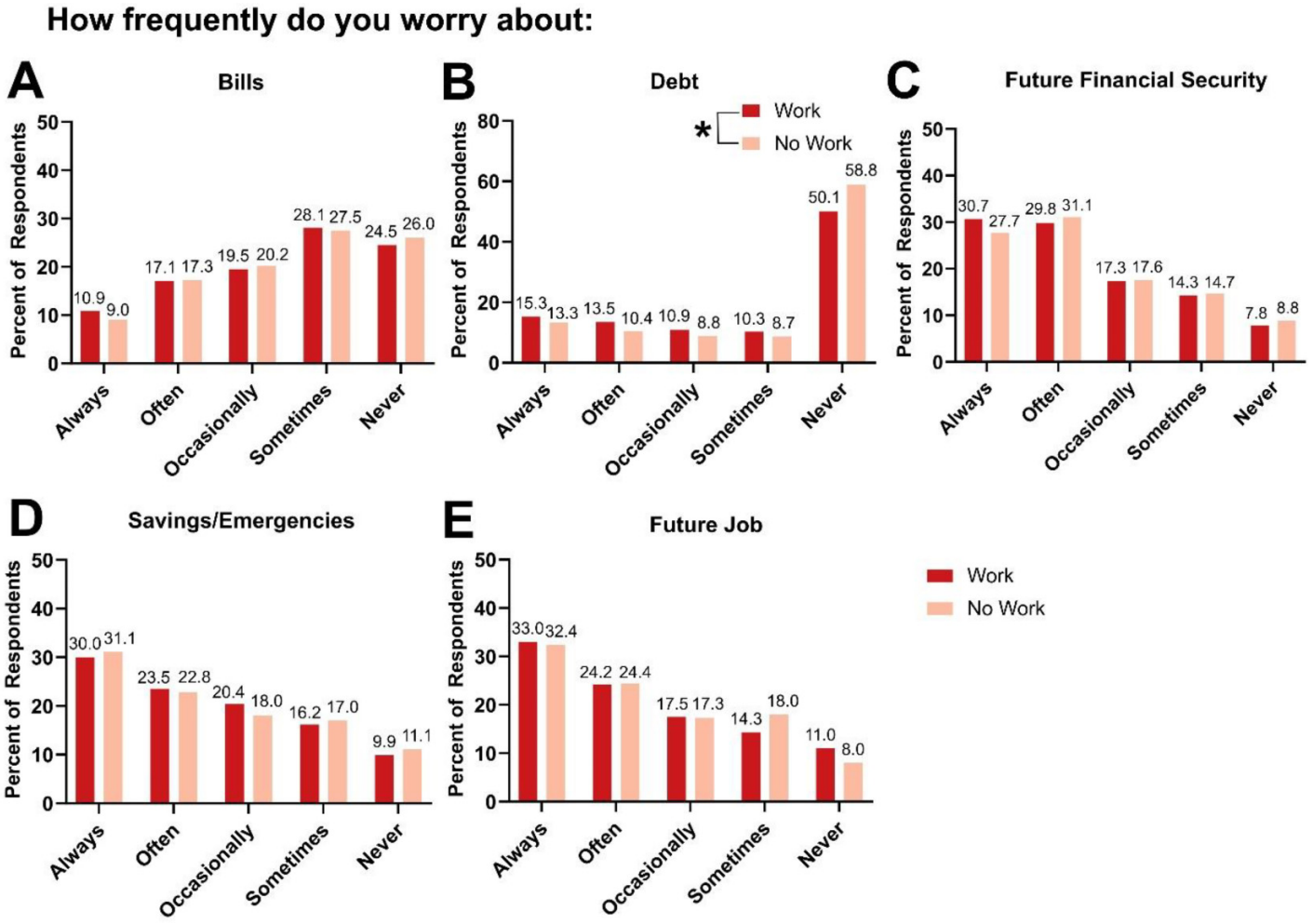
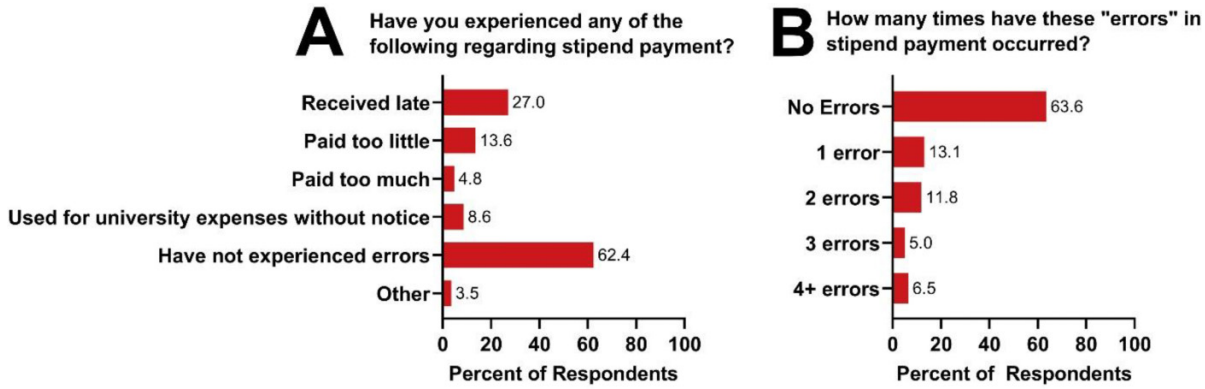


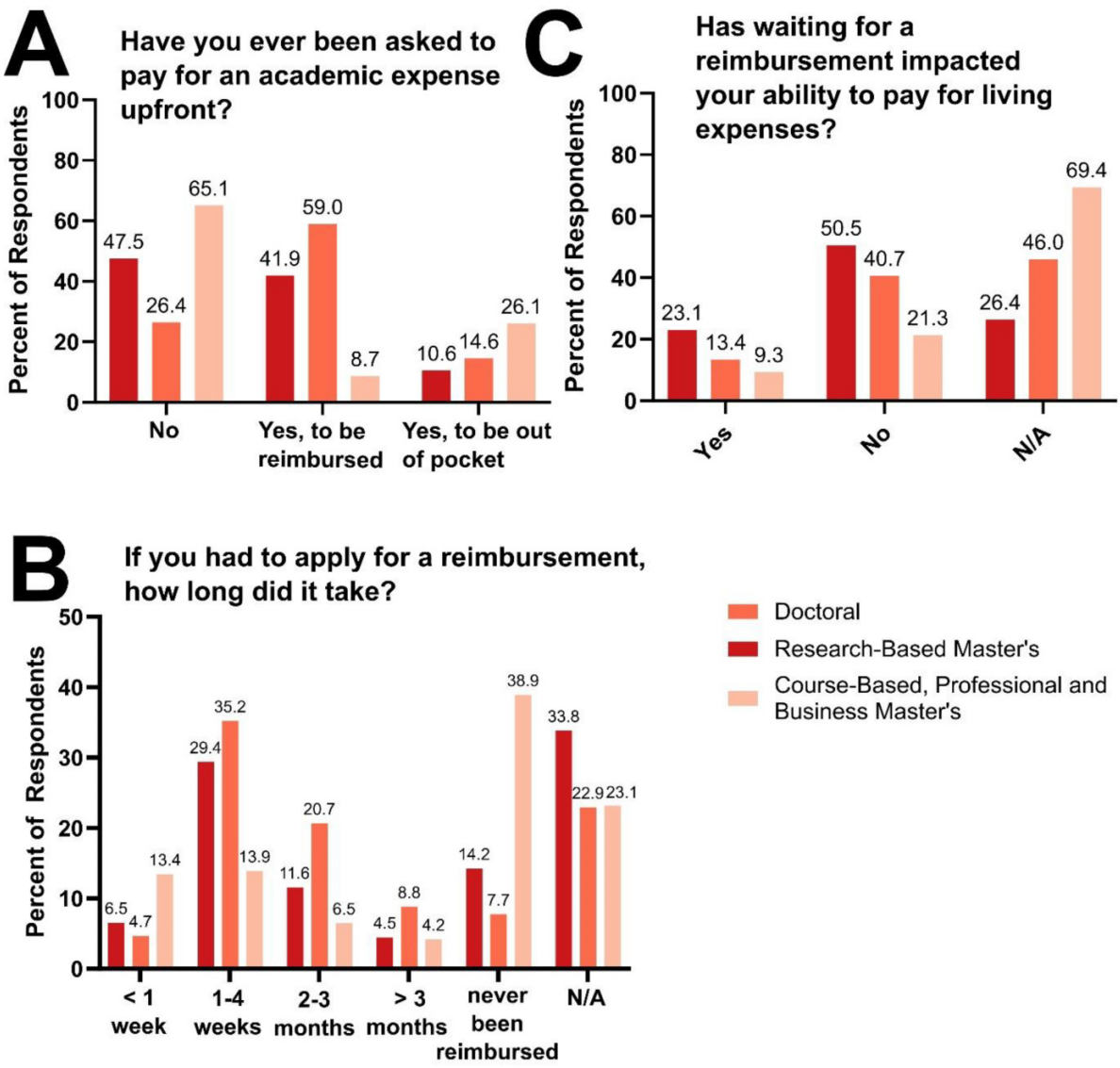
Fig. 16. Stipend errors. (A) Respondents were asked whether they ever experienced any errors regarding stipend payments (multiple selections possible) by % ($n = 914$). An "error" was defined as a *disruption in the regular scheduled payment of your stipend.* (B) Number of times respondents experienced an "error" in stipend payments, by % ($n = 914$).



eral/provincial/institutional scholarships, bursaries/grants, federal/provincial student loans, bank loan/line of credit, family/spousal/partner support, employment pay, employer support, institutional stipends, and personal savings. Though the list of potential monetary support for students may initially appear extensive, the amount of available funding for

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 17. Reimbursements. (A) Respondents indicated whether they have been asked to pay an academic expense upfront, separated by doctoral ($n = 768$), research-based master's ($n = 341$), and course-based, professional, and business master's ($n = 218$) programs. (B) Respondents indicated how long it took to receive a reimbursement, separated by doctoral ($n = 750$), research-based master's ($n = 337$), and course-based, professional, and business master's ($n = 216$) programs: 3–6 month, 6–12 month, and >12 month categories were combined due to too few respondents. (C) Respondents indicated whether waiting for the reimbursement impacted their ability to pay for living expenses, separated by doctoral ($n = 750$), research-based master's ($n = 337$), and course-based, professional, and business ($n = 216$) programs.



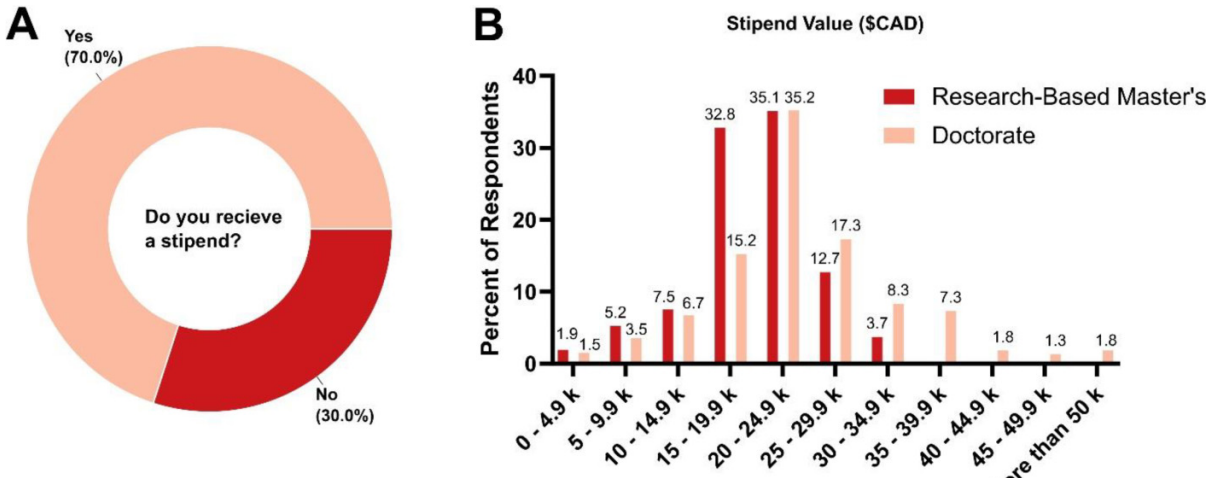
a given student depends on their living situation and varies greatly.

In Canada, awards are granted to students who display a high level of academic achievement and are intended to encourage graduate students to dedicate more time to their studies (Table 2). The awards are primarily distributed by three federal research funding groups: the NSERC, the SSHRC, and the CIHR, which together form the Tri-Councils Agency. In addition to federal research awards, there are also provincial awards for graduate students. For example, the Ontario Graduate Scholarship is allocated through a merit-based program and available for application by students of varying academic disciplines across the province. Each award is jointly funded by the Province of Ontario (66%) and the university offering the award (33%). Fonds de recherche du Québec (FRQ) is another provincial organisation known to recognize excellence in research. Every year, FRQ provides financial support to many students by offering awards, from master's training scholarships to doctoral scholarships. Similar provincial graduate scholarships are offered in BC (BC Gov News 2021), AB (Government of Alberta n.d.), and NS (Communications Nova Scotia 2023).

Of 1305 survey respondents, 17.2% of doctoral and 13.4% of research-based master's students reported receiving a federal award, while 11.2% of doctoral and 5.0% of research-based master's students reported receiving a provincial award

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 18. Stipend status. (A) Respondents were asked whether they received a stipend, by % ($n = 1305$). (B) Stipend value of respondents in research-based master's ($n = 268$) or doctorate ($n = 600$) studies. Data are shown by the % of students in the respective level of study. (C) Average and median stipend value of respondents in research-based master's ($n = 268$), course-based master's ($n = 19$), business or professional master's ($n = 21$), or doctorate ($n = 600$) studies.



C

	Doctorate	Research Master's	Course Master's	Business & Professional
Average	$23,749.68	$19,724.69	$16,812.63	$13,113.19
Standard Deviation	$9,024.50	$6,065.99	$12,144.30	$5,983.74
Median	$22,000.00	$20,000.00	$13,000.00	$13,577.00

Fig. 19. Stipend satisfaction: respondents were asked whether they were happy with the value of their stipend based on the Likert scale of satisfaction (1 as very unhappy and 5 as very happy), by % ($n = 914$).

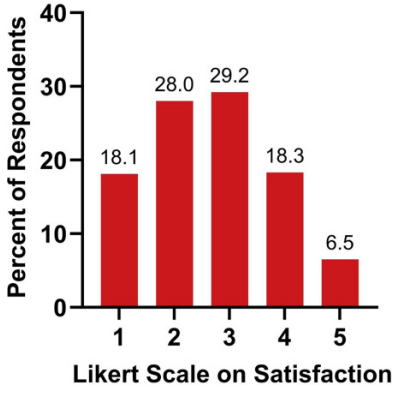
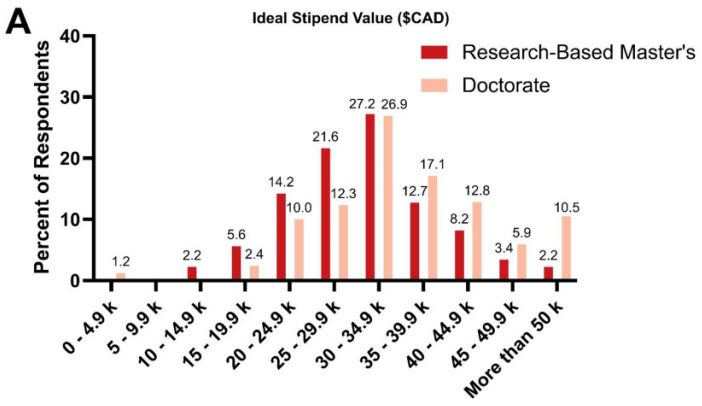


Fig. 20. Ideal stipend. (A) The ideal stipend value shared by respondents in research-based master's ($n = 267$) or doctorate ($n = 592$) studies who responded "yes" to receiving a stipend. Data are shown by the % of students in the respective level of study. (B) Average and median of ideal stipend value of respondents in research-based master's ($n = 267$), course-based master's ($n = 17$), business or professional master's ($n = 21$), or doctorate ($n = 592$) studies.



B

	Doctorate	Research Master's	Course Master's	Business & Professional
Average	$33,619.16	$29,212.93	$26,911.76	$22,704.76
Standard Deviation	$10,525.67	$11,046.15	$12,535.24	$21,661.38
Median	$32,000.00	$30,000.00	$24,000.00	$20,000.00

(Fig. 26). These award recipients represent high-calibre students who have not only achieved acceptance into competitive graduate research programs but have also been identified as the top scholars within this strong research categorization. In surveying graduate students in Canada, we wanted to identify whether the intended goals of federal and provincial awards were being achieved by the current 2021 award values (Canadian Institutes of Health Research 2014) and how the financial well-being of award and non-award holders varied.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 21. Stipend sources. (A) Current sources of stipend of respondents (multiple selections possible) ($n = 914$). Other includes institutional scholarships, research assistantships, and university scholarships. (B) Length of time funded through current sources of stipend ($n = 914$). (C) Previous sources of stipend, not including current stipend sources (multiple selections possible) ($n = 914$). Other includes Mitacs and institutional scholarships. (D) Number of current stipend sources ($n = 914$). Data are shown by percentage.

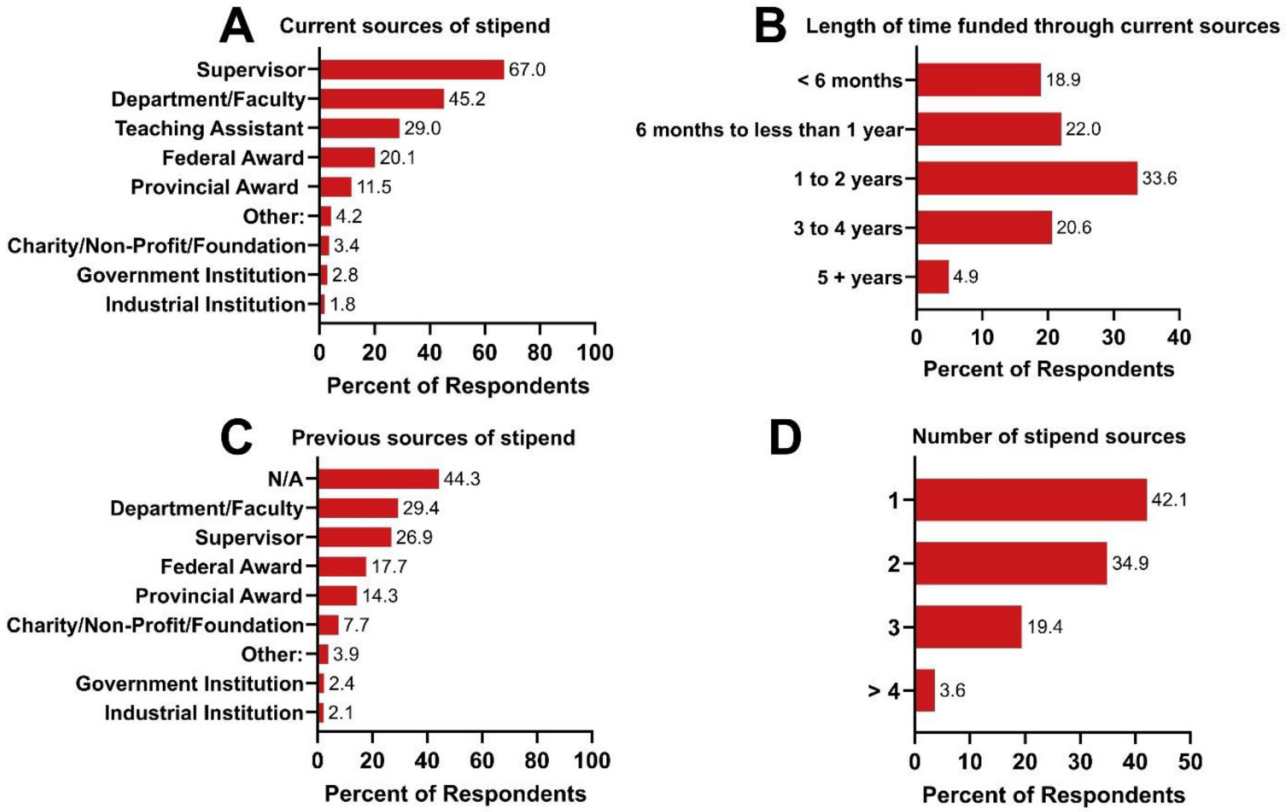
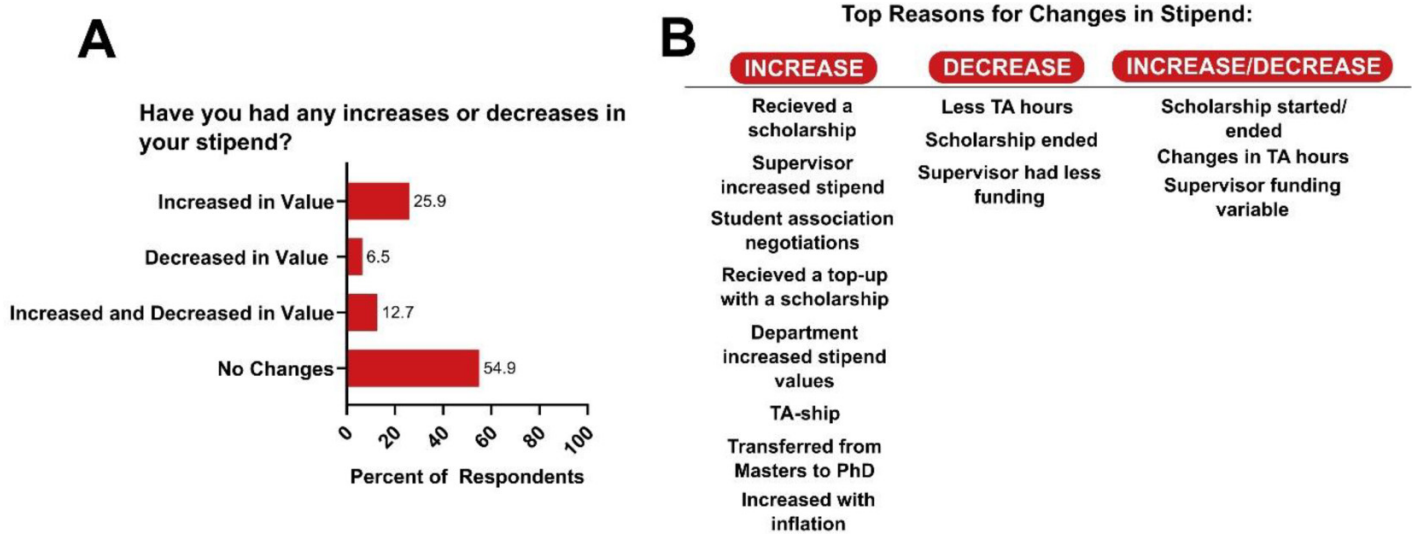


Fig. 22. Stipend changes: (A) reported increases and decreases in student stipend values ($n = 914$) and (B) top reasons for the reported change in stipend values, broken down as increase, decrease, or both an increase and a decrease.



In comparing responses of the financial concerns of award and non-award holders, awardees reported having less financial concern for bills (significant difference at $P < 10^{-5}$ level using a χ^2 test) and savings/emergencies (significant difference at $P = 0.032$ level using a χ^2 test) than non-award holders (Fig. 27). When asked how much savings students have available in case of emergencies, award holders were significantly more likely than non-award holders to have 6+ months of liv-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 23. Tuition fee by field of study. Average annual tuition in CAD by field of study based on survey responses ($n = 1305$). Respondents self-reported their field of study from seven categories. If respondents indicated multiple fields of study or an option not listed in the survey, they were categorized into one of the six categories based on their answers for analysis purposes. Bars represent mean ± SD.

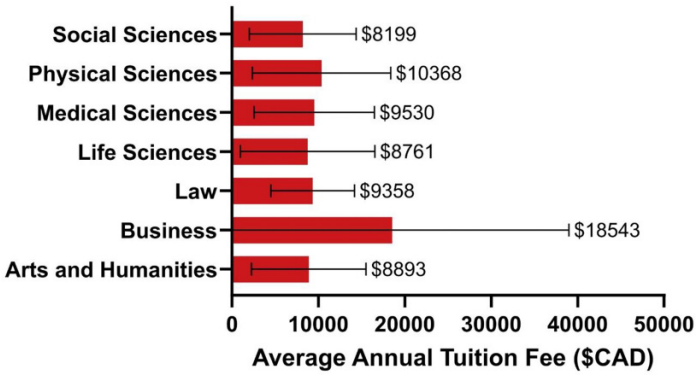
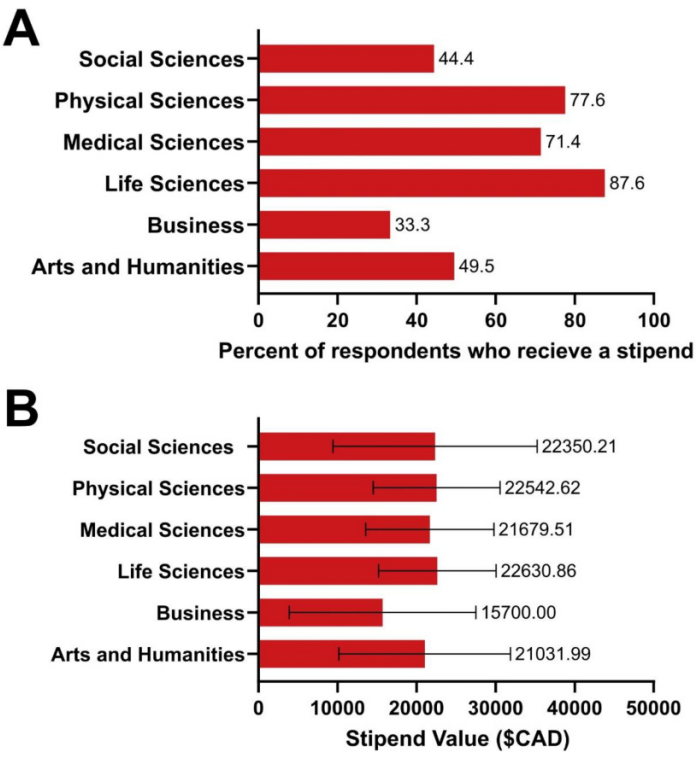


Fig. 24. Field vs. stipend. (A) Respondents were asked whether they receive a stipend, categorized by field of study. Responses are shown as a % of respondents who responded yes to receiving a stipend ($n = 914$). (B) Average stipend value in Canadian dollars (CAD) of respondents by their field of study. Arts and humanities ($n = 101$), business ($n = 21$), life sciences ($n = 307$), medical sciences ($n = 311$), physical sciences (engineering, physics, mathematics, etc.) ($n = 344$), and social sciences ($n = 214$). Law was excluded due to less than five responses. Bars represent mean ± SD (B).



ing expenses saved (χ^2 test, $P = 0.00011$), though we note that less than 30% of respondents in both categories claimed to have these funds available (Fig. 28). The survey revealed that award holders have less outstanding student debt compared to non-award holders, at $14 260 and $17 977, respectively (Fig. 29). Furthermore, when asked to describe their current financial situation, award holders were less than half as likely to respond as "struggling" and more than twice as likely to respond as "comfortable" compared to non-award holders, a significant difference at the $P < 10^{-5}$ level (χ^2 test) (Fig. 30). Overall, 48.4% of non-award holders and 38% of award holders responded as either "struggling" or "tight" to describe their current financial situation. These survey findings are not surprising, considering that awards can help guarantee an income for graduate students that may or may not be provided to non-awardees.

Surprisingly, however, the survey responses demonstrated that awardees are nonetheless preoccupied with financial concerns during their graduate program. For example, when asked how frequently they worry about financial concerns including bills, savings/emergencies, and future financial security, over 66% of award holders responded having financial concern through responses of "always," "often," or "sometimes." Over 44% of award holders responded having concerns for financial debt (Fig. 27). These findings highlight the inability of current federal and provincial award values to effectively enable graduate award holders to fully concentrate on their studies. When graduate students were asked what their ideal stipend value would be, 26% of respondents stated $30 000–$34 999 (Fig. 20). This value range falls close to the value of the current Canadian Graduate Scholarships-Doctoral (CGS-D) award at $35 000 per year.

Additionally, we examined financial incentives received by graduate students after receiving an award (Fig. 31). The most popular incentive was an additional "top-up" of their stipend (27% of respondents), typically from their supervisor. This is typically granted to the student in cases where the scholarship replaces a stipend previously paid by their supervisor. Since the supervisor no longer needs to pay the student's stipend, some will choose to pay the student a smaller additional amount on top of their award. Additionally, 17% of respondents reported receiving a scholarship meant to cover their tuition fully, and 14% received a scholarship meant to cover their tuition partially. This is often granted by the student's department/university in the form of an "Excellence Scholarship" or "Tuition Waiver."

Interestingly, 5.5% of award holders reported a decrease in their total amount of take-home money. This could be due to specific situations where the value of awards is less than the stipend value they were previously receiving. An additional concern is that these awards are defined for a period of time that often does not cover the entirety of their studies. Although not fully captured in this study, future efforts should identify what happens to students' financial situations following the end of their award. It was noted various times in the personalized responses that students often did not have guaranteed funding after their scholarship ended. For exam-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

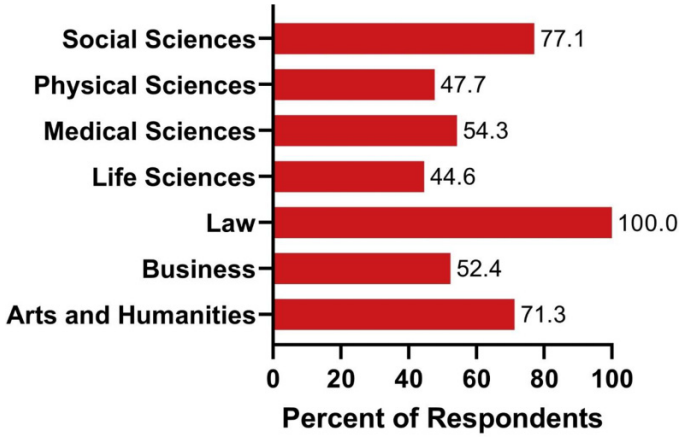
Table 1. Top funding sources by field of study.

Field of study	Ranking	Funding source	Percent of respondents
Arts and humanities	1	Department/faculty	64.00%
	2	Teaching assistant	46.00%
	3	Federal award	22.00%
Life sciences	1	Supervisor	75.09%
	2	Department/faculty	44.61%
	3	Teaching assistant	31.60%
Medical sciences	1	Supervisor	74.77%
	2	Department/faculty	30.63%
	3	Federal award	17.57%
Physical sciences	1	Supervisor	90.99%
	2	Department/faculty	55.86%
	3	Teaching assistant	34.68%
Social sciences	1	Department/faculty	67.37%
	2	Teaching assistant	50.53%
	3	Supervisor	33.68%

Note: Top three funding sources by field of study (multiple selections possible): arts and humanities ($n = 50$), business ($n = 21$), life sciences ($n = 269$), medical sciences ($n = 222$), physical sciences (engineering, physics, mathematics, etc.) ($n = 267$), and social sciences ($n = 95$). Business and law was excluded due to less than five responses. Data are shown as a % of respondents by field of study.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 25. Work by field of study. Respondents were asked whether they work a job outside of their studies, categorized by field of study: arts and humanities ($n = 101$), business ($n = 21$), law ($n = 7$), life sciences ($n = 307$), medical sciences ($n = 311$), physical sciences (engineering, physics, mathematics, etc.) ($n = 344$), and social sciences ($n = 214$). Data are shown as a % of respondents who responded yes to working outside of their studies.



ple, one student comments "For example, [my scholarship] will end in August 2022, and my stipend for my research assistantship will not be adjusted to compensate for the $13 000 net loss in my annual compensation. For myself and others in my position, this will be an extremely large financial burden as we attempt to complete the final years of our graduate studies."

Graduate students from underrepresented groups are more likely to struggle financially

Diversity is essential among graduate students to foster generations of scientists with diverse perspectives and ideas. This will be vital to innovation as we tackle future challenges in society. Per the Dimensions charter: "equity, diversity and inclusion strengthen the research community, the quality, relevance and impact of research, and the opportunities for the full pool of potential participants." (Government of Canada 2019). A 2019 *Survey of Postsecondary Faculty and Researchers* from Statistics Canada showed that postdoctoral fellows and PhD students are increasingly more diverse than the previous generation of researchers. In fact, 50% of postdoctoral fellows and 39% of PhDs who responded to the survey reported belonging to a visible minority group.

Despite the presence and increasing prevalence of diverse groups, women, Indigenous peoples, and visible minorities still face inequities and barriers to inclusion in graduate programs, including lack of representation among those who receive federal doctoral and master's level scholarships (Baskaran et al. 2021). Additionally, it has been found that those who identified as women, visible minorities, individuals with disabilities, and Indigenous were all less likely to receive research funding (Statistics Canada 2020). Factors such as eligibility, duration of awards, and evaluation criteria are significant barriers to promoting equity, diversity, and inclusion (EDI) in graduate programs. While targeted investments, such as new funding in Budget 2022 for Black Scholars (Department of Finance Canada 2022), and recent changes to evaluation criteria, such as NSERC's New Guidelines on Contributions to Research, Training, and Mentoring, which expands the definition of "excellence" (National Sciences and Engineering Research Council of Canada 2022), are steps in

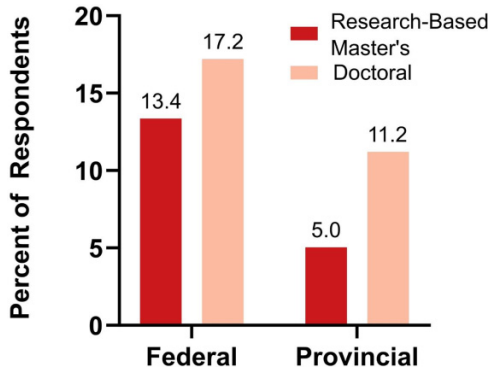
Table 2. Summary of Federal Tri-Council Awards.

Award title	Federal research funding agency	Value per year ($CAD)	Maximum duration	Tenure	Eligibility
Canadian Graduate Scholarships—Master's (CGS-M)	Tri-agency	$17 500	1 year	Eligible Canadian institution	Canadian citizen, permanent resident of Canada, or a protected person under subsection 95(2) of the Immigration and Refugee Protection Act (Canada). Completed between 0–12 months in program
Vanier Canada Graduate Scholarships	Tri-agency	$50 000	3 years	Eligible Canadian institution	Canadian citizen, permanent resident of Canada, foreign citizens. Completed <20 months in program. Completed <32 months in program
Canadian Graduate Scholarships—Doctoral (CGS-D)	Tri-agency	$35 000	3 years	Eligible Canadian institution	Canadian Citizen, permanent resident of Canada, or a protected person under subsection 95(2) of the Immigration and Refugee Protection Act (Canada). Completed <24 months in program. Completed 36 months in program
NSERC Postgraduate Scholarships—Doctoral (PGS-D)	NSERC	$21 000	3 years	Eligible Canadian institution or foreign institution	Canadian citizen, permanent resident of Canada or a protected person under subsection 95(2) of the Immigration and Refugee Protection Act (Canada). Completed <24 months in program
SSHRC Doctoral Fellowships	SSHRC	$20 000	1–4 years	Eligible Canadian institution or foreign institution	Canadian citizen, permanent resident of Canada or a protected person under subsection 95(2) of the Immigration and Refugee Protection Act (Canada). Completed <24 months in program
Doctoral Foreign Study Award (DFSA)	CIHR	$30 000	3 years	Foreign institution	Canadian Citizen, permanent resident of Canada or a protected person under subsection 95(2) of the Immigration and Refugee Protection Act (Canada). Completed <24 months in program

Note: A summary of NSERC, SSHRC, and CIHR awards for master's and PhD students, including the funding agency, value, and eligibility requirements.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 26. Federal and provincial award holders. Percent of respondents who have received an award based on level of study: research-based master's ($n = 337$) and doctoral ($n = 750$). Award-holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.).



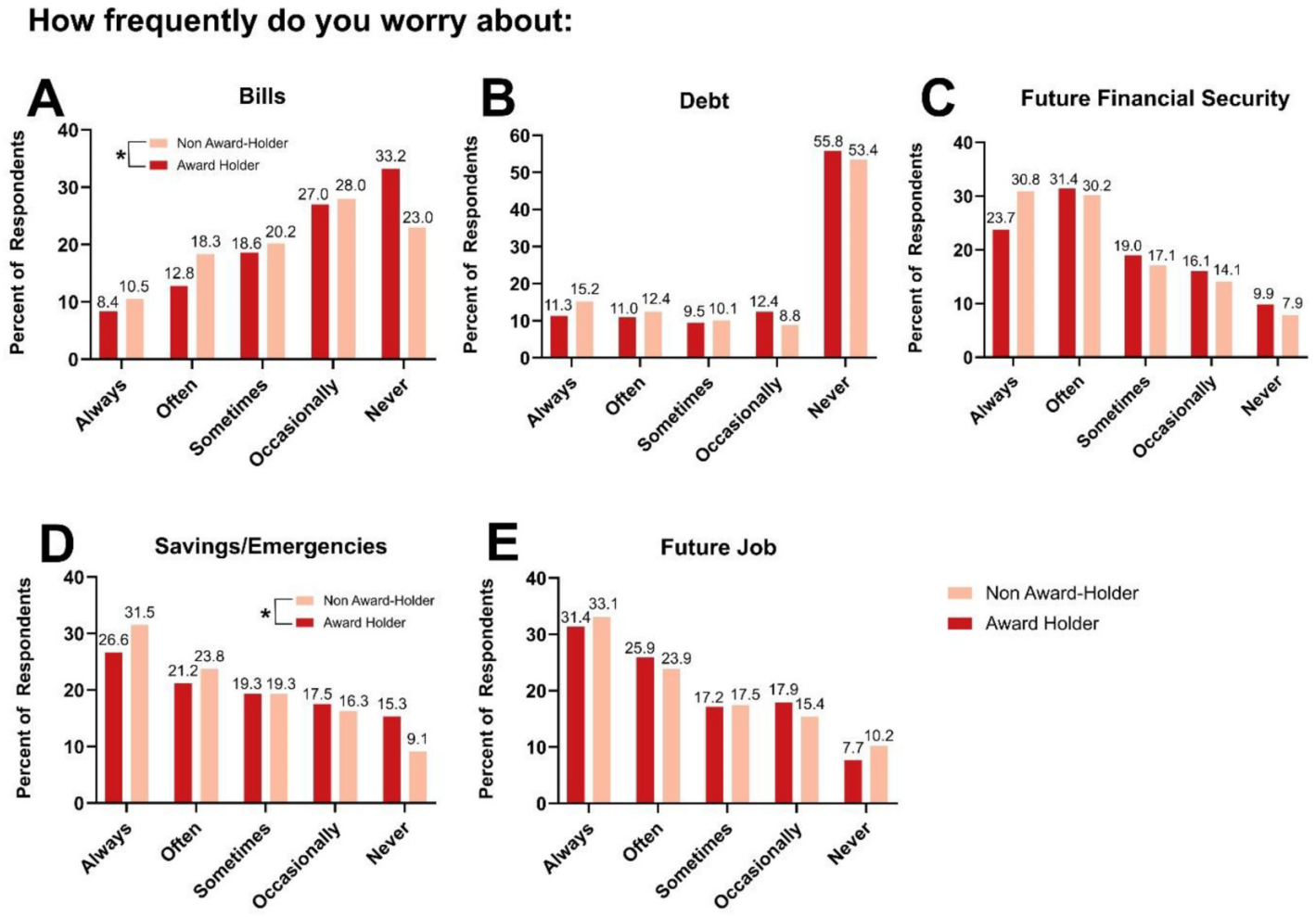
the right direction, there is still work to be done to ensure accessibility of graduate degrees.

This lack of support can lead to a lack of diversity in academia, and this concern was recognized by respondents in our survey. One respondent remarked that they "think that paying students such low stipends (i.e., below the poverty line) is really discouraging EDI in academia and is ensuring that already privileged people are the only ones that have access to higher education." From the data we collected, we aimed to understand the financial struggles of underrepresented groups of graduate students to further identify barriers of entry into graduate degrees in Canada.

Gender-based analysis

Women composed the majority of the survey sample (63.2%), followed by men (32.8%) and respondents who identified as non-binary, gender-fluid, agender, genderqueer, trans, two-spirit, or gender non-conforming (4.0%; category definitions in Supplementary file) (Fig. 32). There were no significant differences in average stipend values among gender groups (KW test, $P = 0.78$) (Fig. 33). Across all gender groups, approximately half viewed their financial situations as "Struggling" or "Tight," with slightly higher numbers among men; however, the differences between groups were not significant (χ^2 test, $P = 0.053$) (Fig. 35). It appears that respondents who identified as either non-binary, gender-fluid, agender, genderqueer, trans, two-spirit, or gender non-

Fig. 27. Financial situation by award holder status. How frequently respondents, based on award holder status, worry about their (A) ability to pay bills, (B) ability to pay back their student debt, (C) future financial security, (D) savings/ability to pay for emergency expenses, and (E) ability to obtain a future job in their chosen field. Data are shown by % in each category: non-award holder ($n = 1031$) and award holder ($n = 274$). Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.). $^{*}P < 0.05$, χ^2 test (award holder vs. non-award holder).



conforming were less likely to have received a government award compared to those who did, though again this difference was not significant (χ^2 test, $P = 0.058$) (Fig. 34).

With the low sample size of gender-diverse individuals in this study, cross-analyses for these metrics were not conclusive. Future efforts should target to better understand the specific financial struggles of these graduate students.

Respondents largely identified as White (66.5%) (Fig. 36). There were no significant differences (χ^2 test, $P = 0.73$) seen in the proportion of domestic students that received a government award (Fig. 38). No significant differences in average stipend values were found for domestic students (KW test, $P = 0.35$) (Fig. 37). There were significant differences (χ^2 test, $P = 0.043$) in the reported level of financial struggle of domestic students with Black, Latin American, Middle Eastern, and South Asian students more likely to report struggling or feeling tight financially than East Asian, Southeast Asian, or White students (Fig. 39). Among international students, there were also significant differences (χ^2 test, $P = 1.7 \times 10^{-4}$), with Middle Eastern, South Asian, and Southeast Asian (not shown in the figure due to small sample size) students reporting financial struggle at a much higher rate. The discrepancy between the financial stress of domestic and international Southeast Asian students may be due to a small sample (with only 17 students and 11 international students).

Indigenous students

The 2021 Census reported that 1.8 million Indigenous peoples reside in Canada. Of these 1.8 million Indigenous individuals, approximately 1.048 million people identify as First Nation, 625 000 identify as Metis, and 70 000 identify as Inuit, representing the three main categories of Indigenous peoples in Canada (Statistics Canada 2022*d*). Indigenous partic-

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 28. Amount of savings by award holder status. Amount of savings, measured by months of living expenses, that respondents have available for emergencies, by award holder status. Data are shown by % in each category: non-award holder ($n = 1031$) and award holder ($n = 274$). Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.). $^{*}P < 0.05$, χ^2 test (award holder vs. non-award holder).

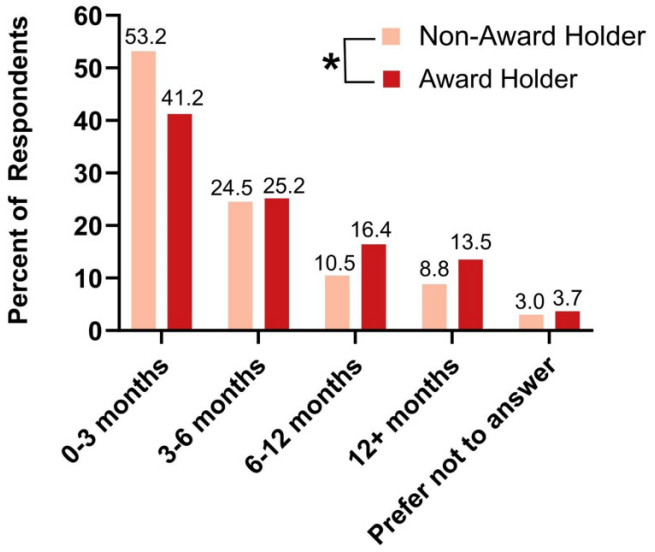


Fig. 29. Average debt by award holder status. Violin plot of average debt of respondents by award holder status. Data are shown by % in each category: non-award holder ($n = 1031$) and award holder ($n = 274$). Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.).The red line and + sign represent the mean and median debt value of respondents in each group, respectively.

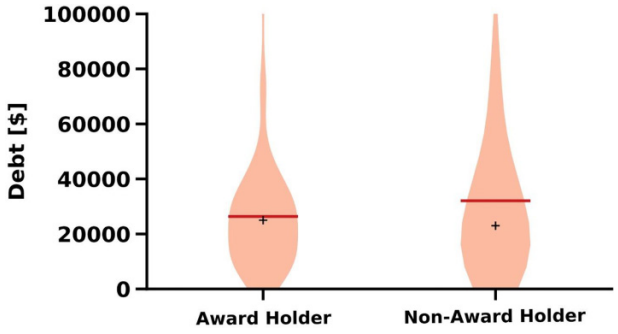


Fig. 30. Financial situation by award holder status. Financial situation of non-award holder ($n = 1031$) and award holder ($n = 274$) respondents. Data are shown by the % of respective award statuses. Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.). $^{*}P < 0.05$, χ^2 test (award holder vs. non-award holder).

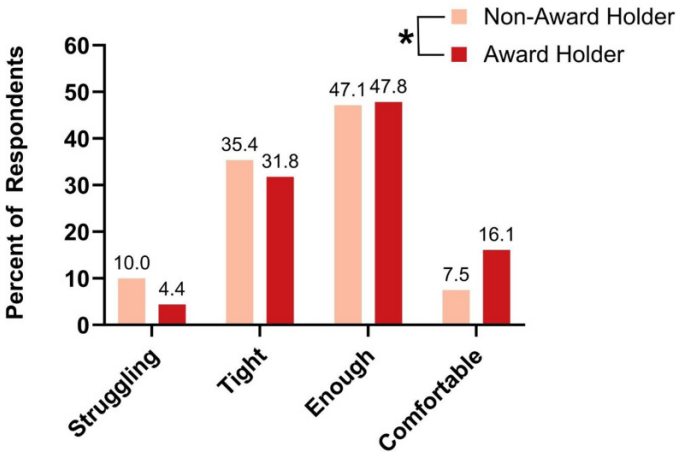
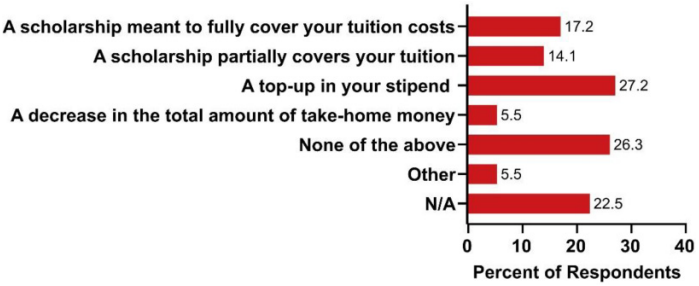


Fig. 31. Scholarship incentives. Respondents were asked whether they received additional incentives or changes in their financial compensation after receiving their scholarships ($n = 941$). Other includes scholarship that is subtracted from stipend (no top up) and funding caps.



ipation in graduate programs surveyed was low. Indigenous respondents consisted of only 1.2% of our survey population, while 0.8% of individuals preferred not to answer this question (Fig. 40). Proportionally, this is consistent with a 2020 Statistics Canada report showing that 1.7% of PhD students identified as indigenous (Statistics Canada 2020). While Indigenous peoples in Canada belong to three broad groups: Inuit, Metis, and First Nations, our survey did not collect any further identifying information among the Indigenous respondents.

Some concerning trends could be seen between Indigenous and non-Indigenous graduate students in the findings. There was a significant reduction (KS test, $P = 0.0012$) in the average stipend value for Indigenous respondents ($13 682) compared to non-Indigenous respondents ($22 286) (Fig. 41). There was no significant difference (χ^2 test, $P = 0.42$) in the responses of Indigenous (33.3%) and non-Indigenous (43.7%) individuals that their financial situation was struggling or tight (Fig. 42). The small sample in our dataset indicates that it could be important for a more detailed investigation into these discrepancies.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 32. Gender identity. Gender identity of all respondents by % ($n = 1305$) (multiple selections possible). Definitions are located in Supplementary file. GNC, gender non-conforming.

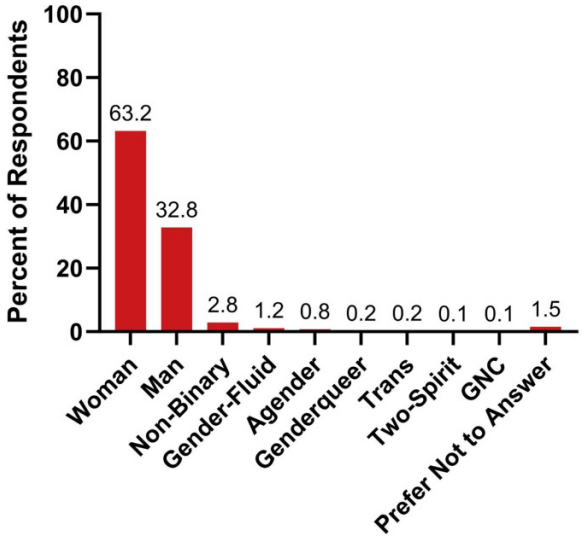
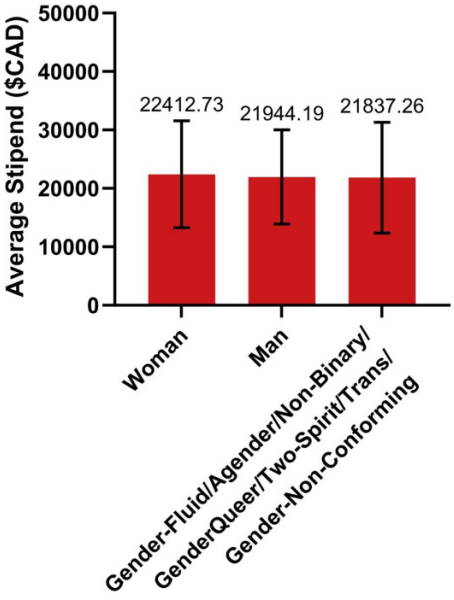


Fig. 33. Average stipend by gender identity. Average stipend value of respondents by their gender identity: woman ($n = 547$), men ($n = 328$), and gender-fluid/agender/non-binary/genderqueer/two-spirit/trans/gender non-conforming ($n = 37$). Bars represent mean ± SD.



Individuals with dependents

A dependent is an individual who relies on others, such as parents and family members, for financial support. We assessed the relevance of graduate students having dependents for their financial struggles. Survey data revealed that the majority of respondents have no dependents (86.5%), while having one dependent (7.4%) or more than one dependent was less common (5.7%) (Fig. 43). In general, having dependents corresponded with increased financial struggle, as more respondents with dependents classified their finan-

Fig. 34. Award holder status by gender identity. Percent of respondents who have received an award by their gender identity: woman ($n = 825$), men ($n = 428$), and gender-fluid/agender/non-binary/genderqueer/two-spirit/trans/gender-non-conforming ($n = 62$). Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.).

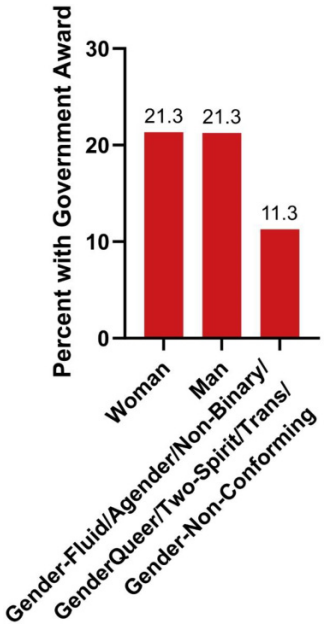


Fig. 35. Financial situation by gender identity. The financial situation of respondents by their gender identity: woman ($n = 825$), men ($n = 428$), and gender-fluid/agender/non-binary/genderqueer/two-spirit/trans/gender non-conforming ($n = 62$). Enough and comfortable responses were grouped to ensure more than five responses in this category. *Ethnicity-based analysis.*

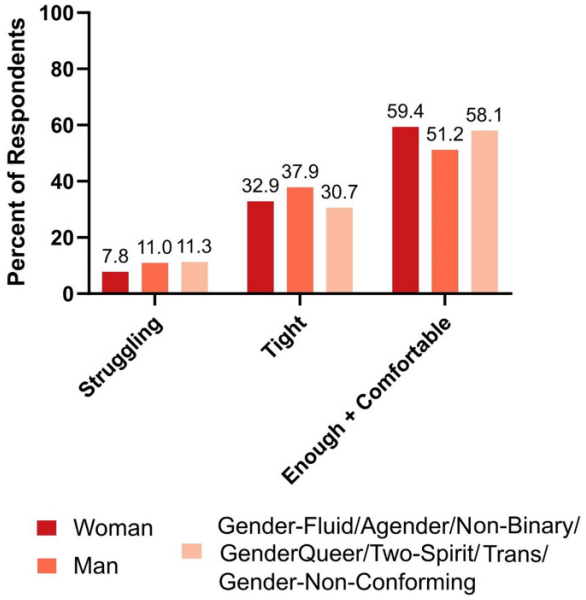


Fig. 36. Ethnic identity. The % of respondents reporting selected ethnic identity (multiple selections possible) ($n = 1300$).

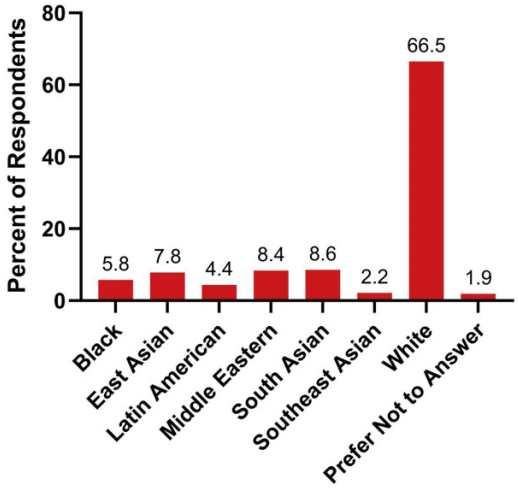
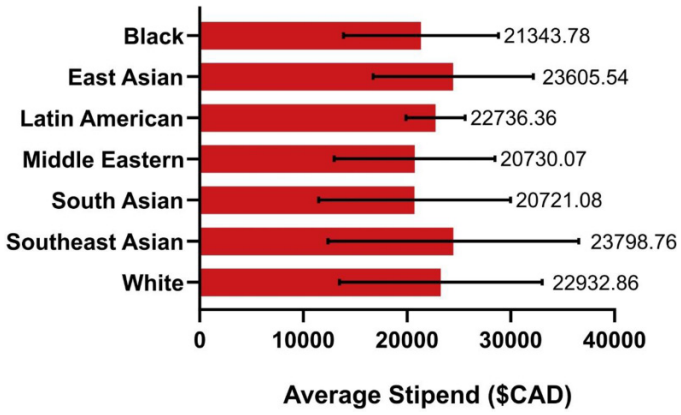


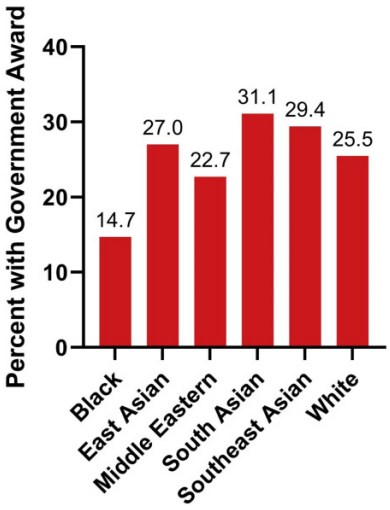
Fig. 37. Average stipend by ethnic identity. Average stipend value of domestic student respondents by their ethnic identity: Black ($n = 23$), East Asian ($n = 54$), Latin American ($n = 11$), Middle Eastern ($n = 29$), South Asian ($n = 36$), Southeast Asian ($n = 11$), and White ($n = 415$).



cial situation as struggling (15%) or tight (42.4%), in comparison to those without dependents (7.8% and 33.6%, respectively) (Fig. 44). This could potentially be a consequence of the struggle of older graduate students with dependents to access need-based funding. One respondent noted that "awards and bursary applications are geared to young students and not at all applicable to adult homeowners of students with families/children. Adult learners are effectively excluded from needs-based support."

Whether a graduate student has dependents or not, they have some contribution to the average debt accumulated in graduate studies. Respondents without dependents accumulated an average debt of $30 889, while those with one dependent averaged $34 933 debt (Fig. 45). As the number of dependents increased, the average debt decreased to below

Fig. 38. Award holder status by ethnic identity. The % of respondents who are domestic students and have received an award by their ethnic identity: Black ($n = 75$), East Asian ($n = 102$), Middle Eastern ($n = 109$), South Asian ($n = 112$), Southeast Asian ($n = 28$), White ($n = 868$). Latin American not shown due to small sample size. Award holder status is defined as being selected to receive a federal or provincial award. This does not include any other awards that were indicated (admission scholarships, institutional scholarships, etc.).



$26 000 (Fig. 45). The majority of graduate students, whether they have dependents or not, have rental accommodations. As the number of dependents increased, there was a slight decrease in the average number of respondents that lived in rental properties and an increase in respondents who owned their homes (Fig. 46A).

We also surveyed respondents on their average monthly housing costs (mortgage or rent). On average, the monthly housing costs tend to be higher for graduates with more dependents, which could be attributed to the requirements of a larger living space. While graduate students with no dependents had an average monthly housing cost of $898, the average monthly housing costs also increased as the number of dependents increased (one dependent—$1204, two dependents—$1439, three dependents—$1034, and four or more dependents $1541) (Fig. 46B).

Individuals living with disabilities

A disability is defined by the Canadian Government as a physical, mental, intellectual, cognitive, sensory, learning, and communication impairment or a functional limitation, whether visible or not and permanent, temporary, or episodic in nature, which hinders a person's full and equal participation in society when they face a barrier. According to the 2017 Canadian Survey on Disability, 22% of the Canadian population aged 15 years or older (approximately 6.2 million individuals) had one or more disabilities (Statistics Canada 2018). It has been well documented that individuals with a

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 39. Financial situation by ethnic identity. (A) The financial situation of domestic student respondents by their ethnic identity: Black ($n = 34$), East Asian ($n = 74$), Latin American ($n = 16$), Middle Eastern ($n = 44$), South Asian ($n = 45$), Southeast Asian ($n = 17$), White ($n = 695$). (B) The financial situation of international student respondents by their ethnic identity: Black ($n = 41$), East Asian ($n = 28$), Latin American ($n = 41$), Middle Eastern ($n = 65$), South Asian ($n = 67$), White ($n = 173$). Southeast Asian excluded due to too few responses. Tight/struggling responses were combined to ensure more than five responses in the grouped categories. $^{*}P < 0.05$, χ^2 test.

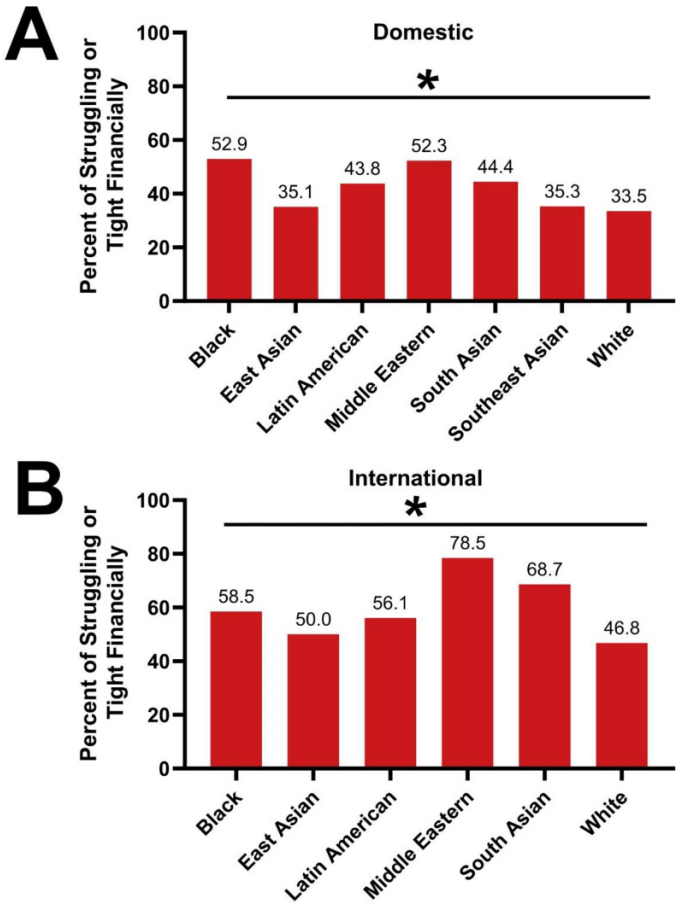


Fig. 40. Indigenous identity. The % of respondents who identified as Indigenous (First Nations, Métis, or Inuit) ($n = 1305$).

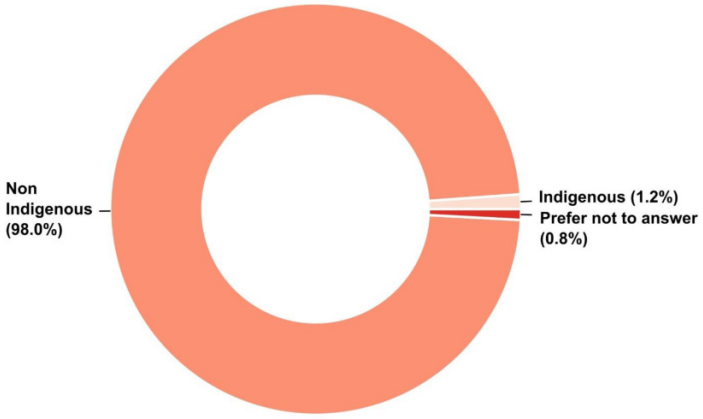


Fig. 41. Average stipend by Indigenous identity. Average stipend value of respondents who identify as Indigenous: Indigenous ($n = 11$) and non-Indigenous ($n = 890$). Bars represent mean ± SD.

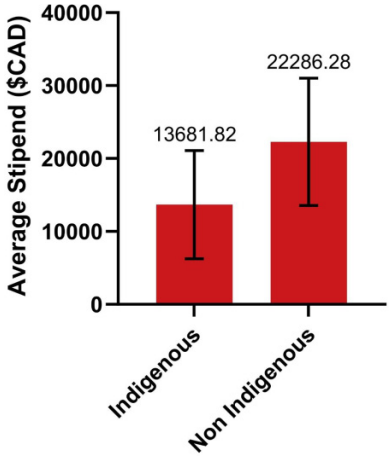
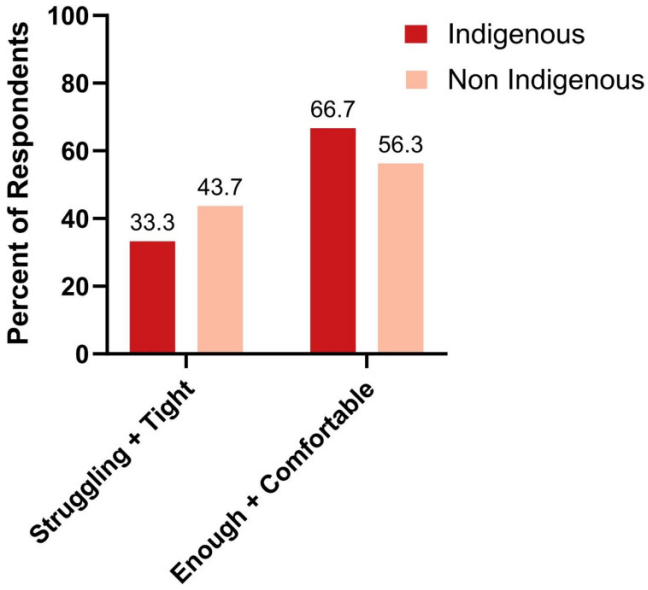


Fig. 42. Financial situation by Indigenous identity. Financial situation of respondents who identify as Indigenous: Indigenous ($n = 15$) and non-Indigenous ($n = 1279$). Enough/comfortable and tight/struggling responses were combined to ensure more than five responses in the grouped categories.



disability face numerous obstacles and challenges in society. Some government aid programs for students with disabilities exist. For example, the Canada Student Grant for Students with Disabilities enables students to apply for grants up to $4000 per academic year through provincial granting agencies (Employment and Social Development Canada 2022*a*). In addition, local funding opportunities exist with grants and

Fig. 43. Number of dependents. The % of respondents with each number of dependents ($n = 1305$).

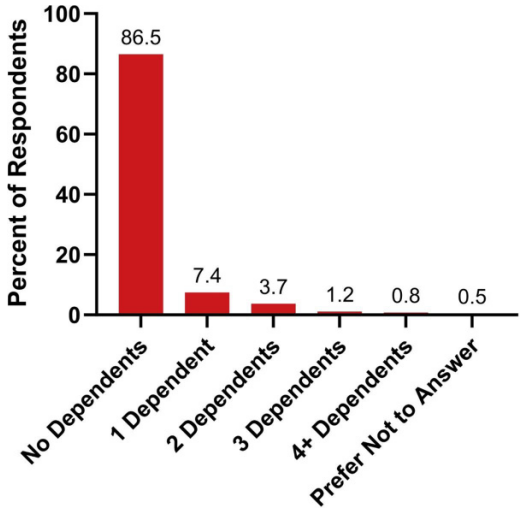
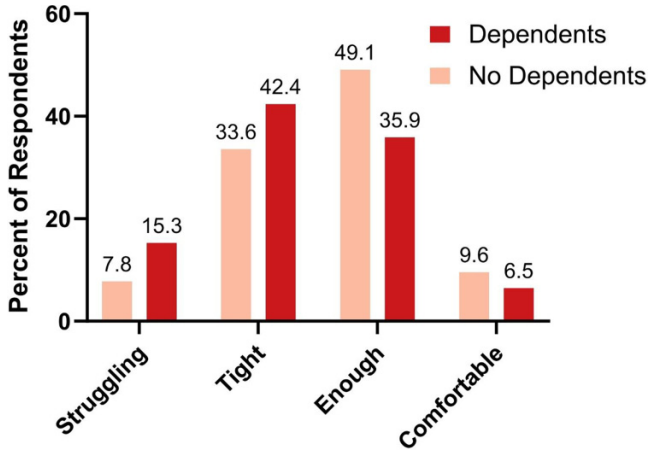


Fig. 44. Dependents and financial situation. The financial situation of respondents with or without dependents: dependents ($n = 170$) and no dependents ($n = 1129$). Dependents includes one or more dependents and were combined to ensure more than five responses.



bursaries for individuals with disabilities at each academic institution. In addition to government funding, the National Education Association of Disabled Students (NEADS) is an organisation founded in 1986 designed to support full access to education and employment for post-secondary students and graduates across Canada. NEADS offers disabled students services such as student debt reduction, increased campus involvement, experience, and help with postgraduate employment. Academic institutions can provide and accommodate services for individuals with disabilities through means such as note-taking, visual language interpretation, captioning for virtual learning, accessibility in campus/classroom, testing/exam accommodation, and tutoring services designed for each disability and learning requirement.

We collected data on whether respondents identified as having a disability and whether it impacted their graduate school success and financial status. Overall, the majority of respondents (87.4%) stated that they did not identify with a disability, 10.3% identified as having a disability, and 2.3% of respondents preferred not to answer this question (Fig. 47). Respondents with a disability had an average stipend value similar to those without a disability ($23 134.83 vs. $22 120.41, respectively) (Fig. 48). In addition, graduate students with a disability received an award/scholarship at similar proportions individuals without a disability, 21.5% and 20.9%, respectively (Fig. 49). While not significant, a larger percentage of students with a disability indicated that they are financially "struggling" or "tight" compared to those without a disability. Our survey reported that 10.4% of graduate students with a disability were "struggling," while only 8.4% of graduate students without a disability were struggling (Fig. 50). In addition, 38.5% of graduate students with a disability were feeling "tight" on their finances, while only 34.0% of graduate students without a disability responded similarly (Fig. 50). Results showed that 51.1% of respondents with a disability were comfortable with their financial status, while 57.4% of respondents without a disability were comfortable with their finances. Respondents noted that things like "counselling, physical therapy, and prescription medications are both not accounted for in stipend values and not covered by public insurance here in Canada," which can disproportionately affect graduate students with disabilities. Another student noted that "disability also increases costs by extending the length of people's programs, costing them many thousands of dollars in tuition and fees." Certainly, academic institutions have made progress with maintaining accessibility and offering support/services, but our survey highlights that individuals with disabilities are struggling or are tight with their finances, thus identifying a need for more governmental support.

International graduate students face larger financial struggles than domestic students

International students composed 21% of master's and 37% of doctoral students in Canada, with this proportion rising to nearly 50% for those studying in STEM (Statistics Canada 2022*e*). This is comparable to the 31.6% of survey respondents to our survey who were international students (Supplementary Table 1).

International students can face additional challenges compared to domestic students, such as legal limits on the number of hours they can work off campus, which has recently been suspended until the end of 2023 (Immigration, Refugees and Citizenship Canada 2022), ineligibility for most Canadian government scholarships, as well as other struggles. One notable difference between international and domestic graduate students is the amount they pay in tuition, with 39% of international students reporting annual fees over $15 000 compared to only 3% of domestic students (Fig. 51). While 28% of international students have their tuition entirely covered by additional scholarships or higher stipends, over 66% must pay for at least part of their tuition out of pocket (Fig. 52).

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 45. Dependents and debt. (A) Violin plot of the average student debt for those respondents who have student debt conditional on the number of dependents ($n = 577$). The red line and + sign represent the mean and median stipend value of respondents in each group, respectively. (B) The proportion of respondents who have student debt conditional on the number of dependents ($n = 1305$).

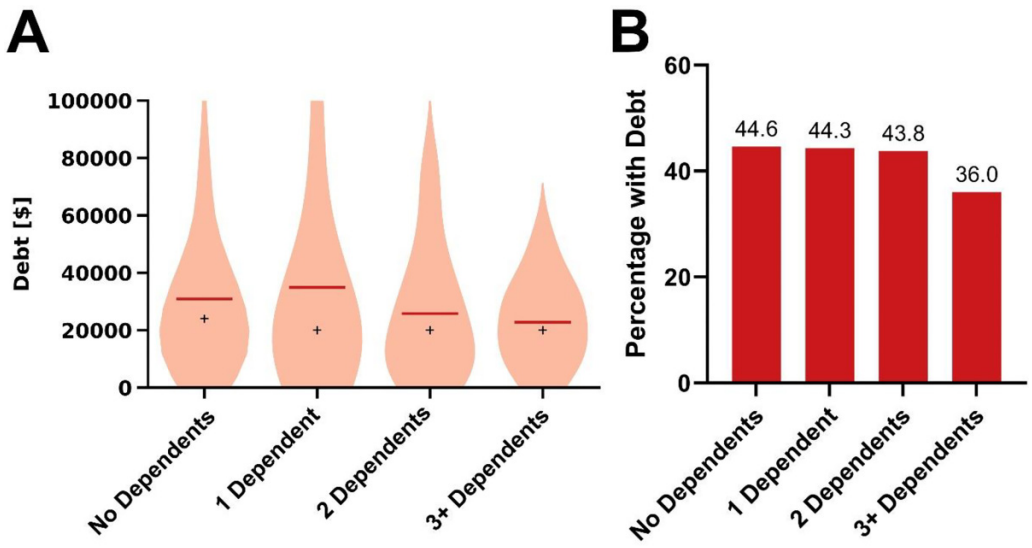
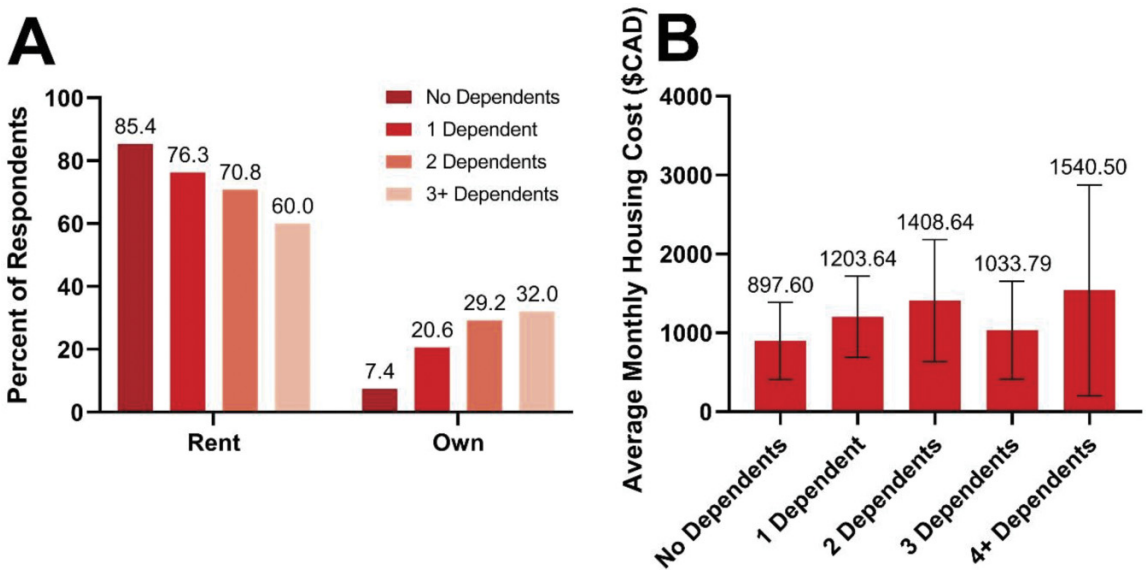


Fig. 46. Dependents and housing. (A) The percentage of respondents who rent or own (does not add up to 100% due to other living statuses) conditional on the number of dependents they have ($n = 1305$). (B) The average monthly cost of housing for respondents conditional on the number of dependents ($n = 1283$). Bars represent mean ± SD.



International students were found to be significantly (χ^2 test, $P < 10^{-5}$) more likely to report that they are "struggling" to make ends meet or that their finances are "tight," with over 50% falling into these categories (Fig. 53). Concerns over one's ability to pay for bills, financial security, and emergency expenses were also significantly higher (χ^2 test, P values of $<10^{-5}$, 0.002, and $<10^{-5}$, respectively). The survey reported 44% of international students were "always worried" about their ability to pay for emergency expenses compared to 24% of domestic students (Fig. 54). The amount of emergency savings held by international students was also significantly different (χ^2 test, $P < 10^{-5}$), with 70% of international students having less than three months of expenses saved compared to 44% of domestic students (Fig. 55). This is despite international students generally paying less in rent (Fig. 56).

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 47. Disability status. The percentage of respondents who identify with a disability ($n = 1305$).

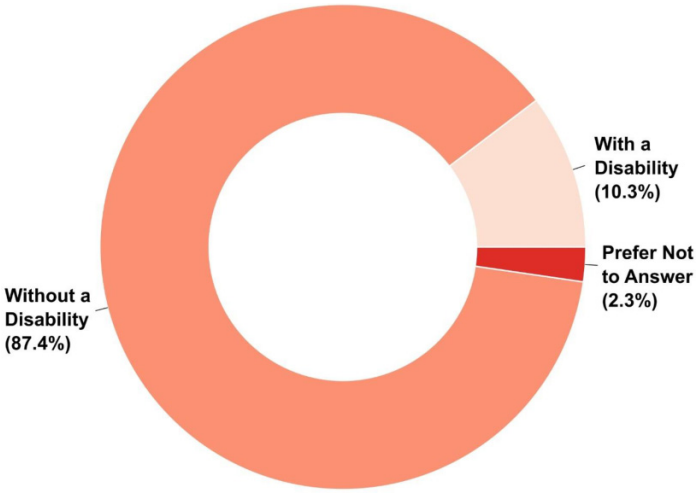
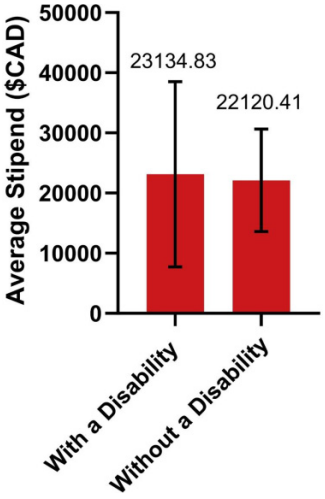


Fig. 48. Average stipend by disability status. The average stipend in Canadian dollars (CAD) of respondents with ($n = 93$) or without ($n = 794$) a disability. Bars represent mean ± SD.



Additionally, there were large differences in how international graduate students received funding. Only 13% of international students received government scholarships compared to 26% of domestic students (Fig. 57). This is likely due to the lack of eligibility of international students for many federal and provincial scholarships. In fact, the only federal scholarship that international students can apply for is the prestigious Vanier Scholarship. An evaluation report of the Vanier Scholarship shows that only 13% of Vanier awardees are international scholars (Canadian Institutes of Health Research 2014). Since these awards, valued at $50 000, are much higher than other federal awards, they are highly competitive and have since failed at their original objective of attracting and recruiting prestigious students from outside of Canada (Canadian Institutes of Health Research 2014).

Fig. 49. Award holder status by disability status. The percentage of respondents with ($n = 135$) or without ($n = 1140$) a disability who indicated receiving a government award.

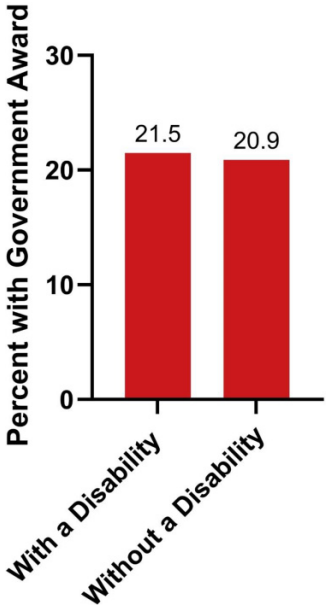
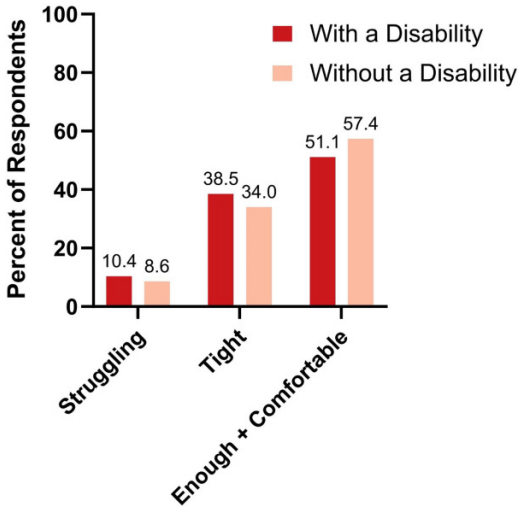


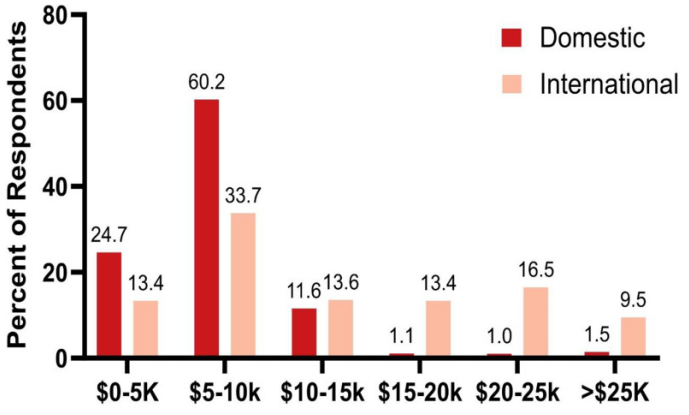
Fig. 50. Financial struggle by disability status. The percentage of respondents with ($n = 135$) or without ($n = 1140$) a disability who identified their financial status as being either struggling, tight, or enough + comfortable. Enough/comfortable responses were combined to ensure more than five responses in the grouped categories.



Further, at the time this survey was conducted, there were limitations on the number of hours international students were allowed to work outside of their studies. This hinders the ability for students to supplement their income and puts them at increasing risk of financial insecurity. International student respondents worked less outside of their studies compared to domestic students (Fig. 58). Recently these limits

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 51. Tuition fees. The annual tuition and other mandatory fees separated by domestic ($n = 892$) and international ($n = 412$) students.



have been suspended (Immigration, Refugees and Citizenship Canada 2022). The restriction for international students to work less than 20 h per week during school semesters has been temporary removed to allow for more additional income and to supplement labour shortages. Similarly, far fewer international student respondents indicated receiving stipends larger than $27 000 compared to domestic students, which may be due to the lack of access to federal scholarships (Fig. 59).

Next steps and recommendations

1. **Increase federal funding (through the Tri-Councils) for graduate students in the form of research grants and scholarships**

The conclusions of this survey support the current recommendation of the Support Our Science campaign (Support our Science 2022). Put simply, we advocate for an increase in the *value* and *number* of graduate student scholarships at academic institutions across Canada to account for increased inflation and enrolment into graduate programs since 2003 (i.e., the last time these awards were increased):

a) Increase in the *value* of Tri-Council scholarships—we recommend harmonising all federal Tri-Council scholarships to a *minimum* of $25 000 for master's students and $35 000 for PhD students. This aligns with the ideal stipend values declared by our survey respondents (Fig. 19). Careful consideration should go into this decision as it can set the standard for stipends at the university level.

b) Increase in the *number* of Tri-Council scholarships—our survey has found that about 20% of students are funded through federal scholarships, while 12% are funded through provincial scholarships (Fig. 21). We recommend increasing the number of students who are funded through these scholarships by 50% annually to account for increased enrolment into graduate programs (Statistics Canada 2022*f*). While the number of graduate students in Canada has doubled between 2002 and 2019, we have seen decreases in the number of Tri-Council scholarships available resulting in increased competition and lower success rates for these competitions. By increasing the number of these awards, it would increase the overall percentage of graduate students funded and remove some financial burden from supervisors and principal investigators to fund these students out of their grant funding.

c) Increase the value of Tri-Council *research grants*—we recommend a specific increase in research grants (e.g., discovery grants, etc.) that will allow for supervisors to increase the stipends of graduate students. Our survey shows that the majority of students (67%) are funded through their supervisors (Fig. 21). Future work should investigate the structural and policy changes needed to ensure that the increases in monetary value of research grants are directly funding and supporting graduate students. Careful consideration should be taken to ensure that financial support for students is increased and not simply being put towards funding for additional students at the current low pay rates.

2. **Implement a standard for stipends in Canada and establish transparency practices for university departments**

Many survey respondents cited confusion in variation and inconsistencies in aspects of stipends. In many ways, this confusion has contributed to the challenges faced by groups advocating for stipend increases. From university and department websites, it is often unclear what funding packages are available for graduate students, and this information can often be buried under policy jargon in academic regulations. Transparency of this information is vital to allow students to make informed decisions about their graduate school pursuits.

Additionally, we recommend developing a standardised approach to defining a stipend for graduate students. This includes shifting terminology to more effectively communicate the "take-home" value of stipends to ensure that students can live above the poverty line. Specifically, this means the amount of a students' stipend that remains *after* paying tuition and mandatory expenses. This distinction is important as tuition remains one of the largest costs for students, and students who pay tuition out of pocket have increased financial struggles (Fig. 7). This goes in hand with increasing transparency of stipends, including stipulations, requirements, timelines, and limitations for graduate students.

As such, we propose the following calculation for universities to establish a minimum stipend for their students:

$$\underline{\text{Minimum stipend}} = \text{Tuition} + \text{average cost of rent} + \$550/\text{month for additional fees} + \$500/\text{month for a PhD student}$$

In this case, tuition would include any ancillary, program, or incidental fees associated with the student's tuition (i.e., all cumulative mandatory tuition fees for the student). This

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 52. Tuition paid out of pocket. The percentage of tuition paid out of pocket separated by domestic ($n = 886$) and international ($n = 406$) students.

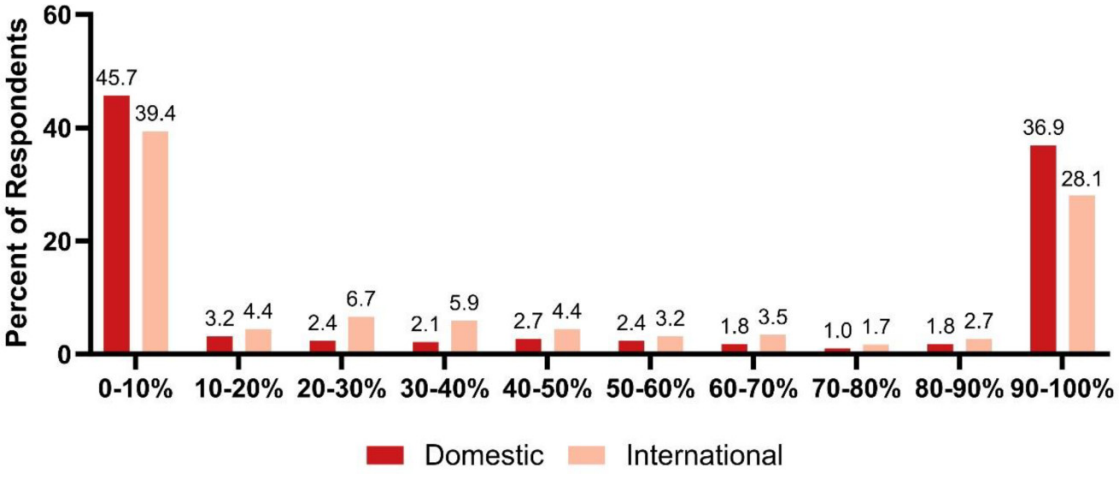
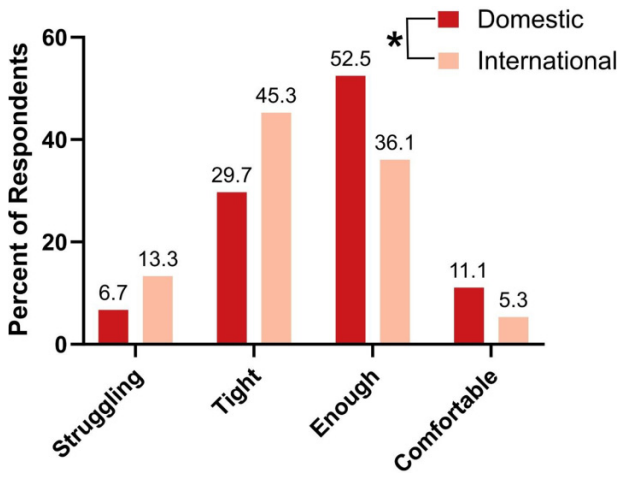


Fig. 53. Financial situation of domestic and international students. The % of respondents who are struggling financially and often do not have enough to make ends meet (struggling); every month is tight and often making sacrifices to pay for necessities (tight); have enough and are able to afford to provide for themselves (enough); or very comfortable and can spend as they like (comfortable), separated by domestic ($n = 892$) and international ($n = 413$) students. $^{*}P < 0.05$, χ^2 test (domestic vs. international).



would also account for an increased stipend need for international students and would compensate if the department provided any tuition waivers for international students. The average cost of rent can be calculated for a one-bedroom apartment within the city where the university is located. We recommend using a resource that regularly publishes these data (such as Zumper: https://www.zumper.com/blog/rental-price-data-canada/) to get the most accurate up-to-date data. Additionally, we recommend an average of $550/month for additional living expenses such as groceries, transportation, medical expenses, and more. The value of this number was calculated in our survey as the average additional monthly expenses seen from students (Fig. 9). We then recommend an additional $500/month for a PhD student as these students have increased financial needs and often have more dependents and responsibilities.

Here, we have provided an example of this equation at work for a domestic (in province, on campus) thesis-based graduate student attending a U15 university in Canada (Table 3). Full details on these numbers and sources can be found in Supplementary files.

3. Expand the eligibility of Tri-Council graduate student scholarships to help provide better funding for international students and underrepresented groups

Current eligibility for Tri-Council graduate student scholarships, namely the CGS and PGS awards (Table 2), excludes international students. While the Vanier scholarship aims to attract and retain international scholars, this award fails to achieve these goals (Canadian Institutes of Health Research 2014). We recommend that the Tri-Council re-evaluate the key priorities of the Vanier scholarship and consider creating separate scholarship allocations for international students as our data indicate that they are less likely to receive these scholarships (Figs. 34 and 38). We further amplify the recommendations of Baskaran et al. (2021), which suggest harmonising the duration eligibility for Tri-Council CGS and PGS awards to increase the time in which a student is eligible to apply for the awards. We also recommend continuing efforts to expand the definition of "excellence" to include diverse forms of work, volunteering, and mentoring. All of these recommendations will assist with removing barriers of entry for foreign scholars and underrepresented groups.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 54. Financial stresses of domestic and international students. The response for domestic ($n = 892$) and international ($n = 413$) students on how frequently they worry about their (A) ability to pay bills, (B) ability to pay back their student debt, (C) future financial security, (D) savings/ability to pay for emergency expenses, and (E) ability to obtain a future job in their chosen field. Data are shown by % in each category. $^{*}P < 0.05$, χ^2 test (domestic vs. international).

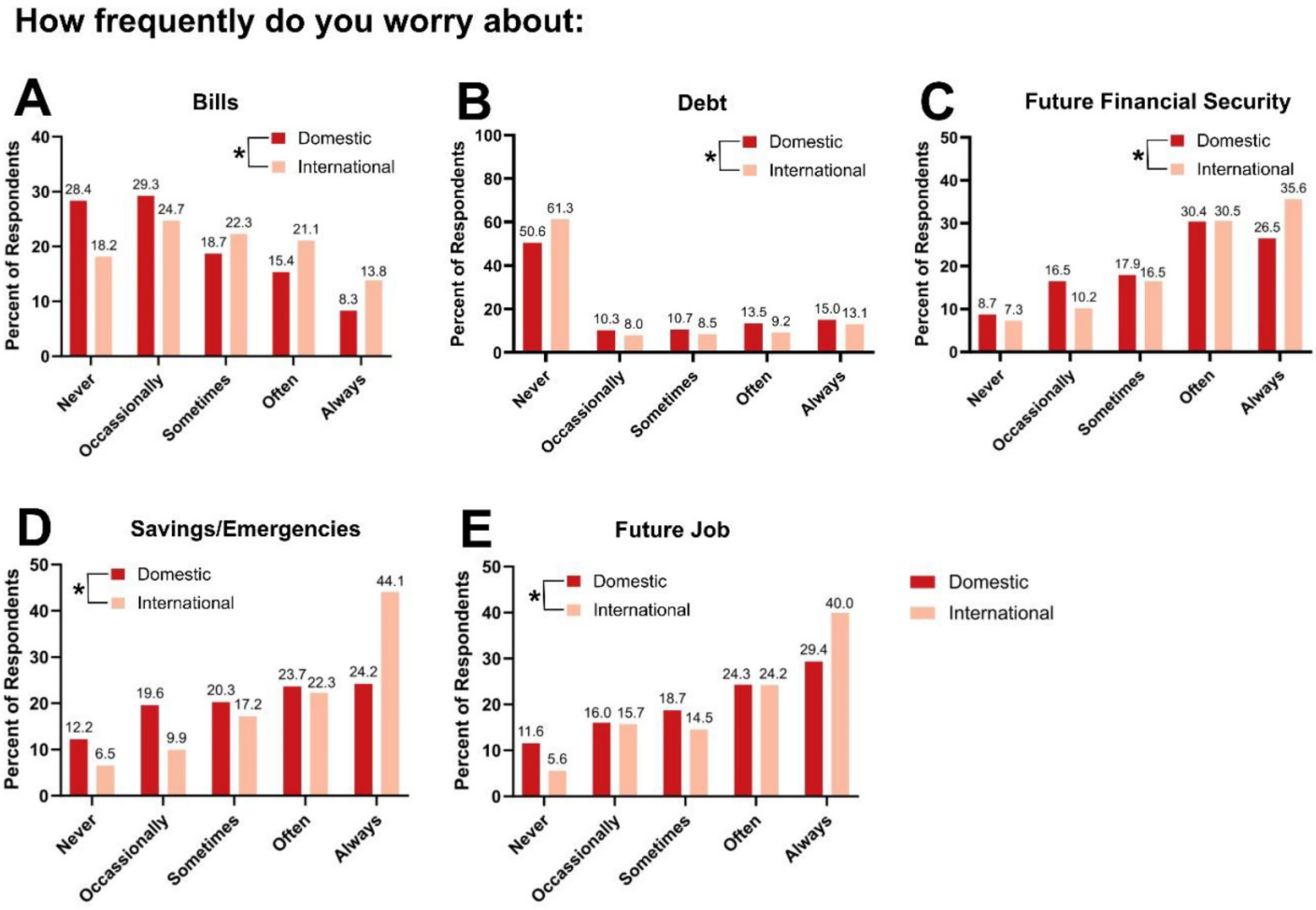


Fig. 55. Savings of domestic and international students. The number of months of living expenses respondents have as savings separated by domestic ($n = 863$) and international ($n = 401$) students. $^{*}P < 0.05$, χ^2 test (domestic vs. international).

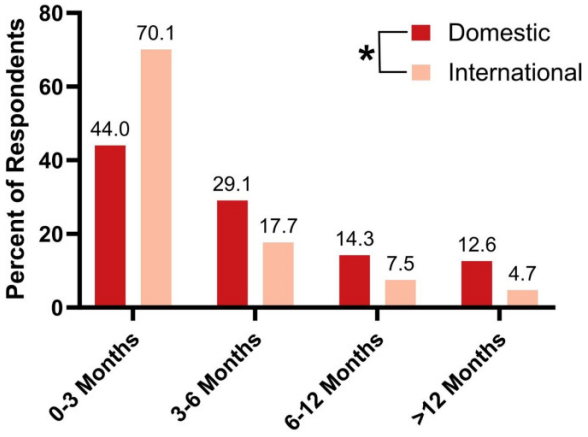
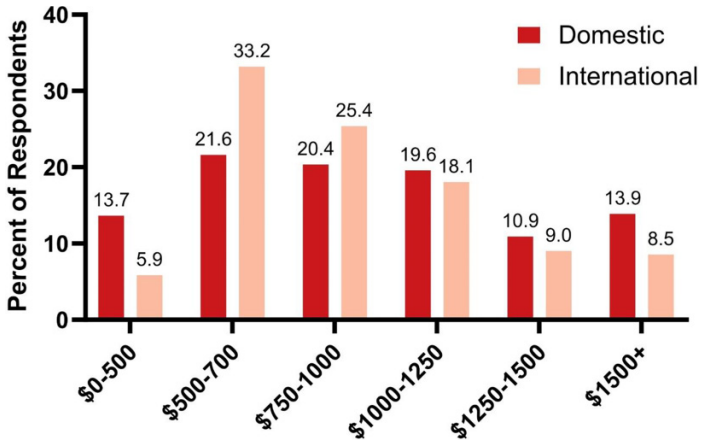


Fig. 56. Housing costs of domestic and international students. The monthly cost of housing of respondents separated by domestic ($n = 879$) and international ($n = 410$) students.



Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Fig. 57. Government awards and leaving studies. (A) The percentage of respondents who receive a federal or provincial award for domestic ($n = 982$) and international ($n = 413$) students. (B) The percentage of respondents who have considered leaving studies due to concern over finances separated by domestic ($n = 889$) and international ($n = 408$) students.

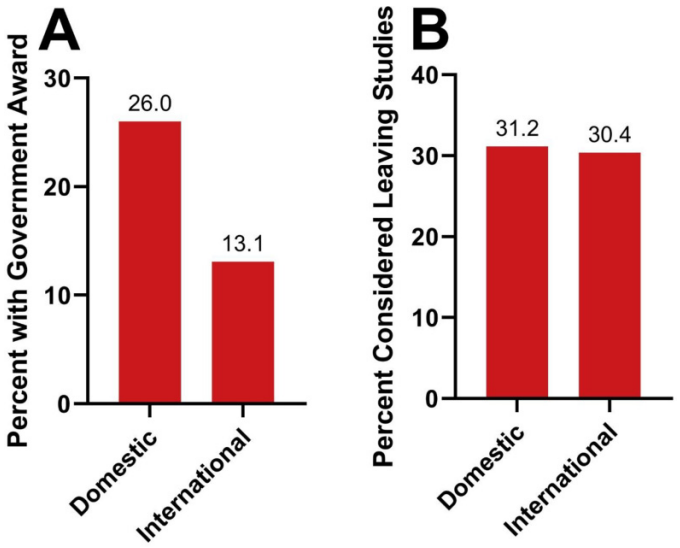


Fig. 58. Hours outside work of domestic and international students. The number of hours worked per week outside of graduate studies separated by domestic ($n = 892$) and international ($n = 413$) students.

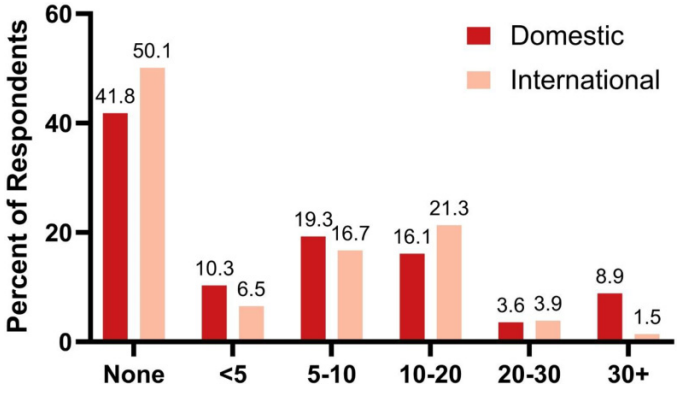
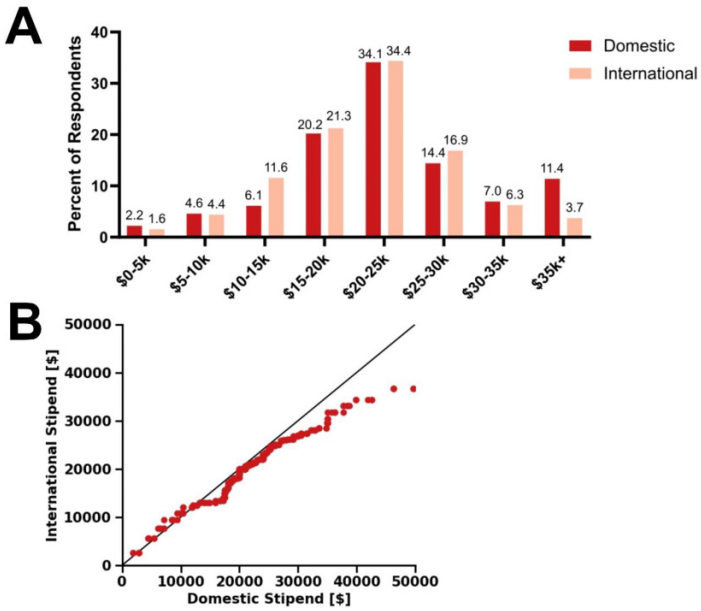


Fig. 59. Stipend levels of domestic and international students. The annual stipend received by domestic ($n = 589$) vs. international ($n = 320$) student respondents. (A) The percentage of respondents in each category who receive a stipend of this amount. (B) A QQ-plot (binned into groups of five to maintain de-identification) showing how the distribution of stipends for domestic and international students differ. A QQ-plot is such that the same quantile for each data set is plotted against each other (e.g., the 5th percentile of both distributions is one point). If the two distributions were the same they would follow the line indicated.



such as that we are experiencing now. The costs of living have continued to increase over time, including the cost of tuition for graduate students, which has increased by 38% since 2006 (Statistics Canada 2022*a*). It is therefore important to recognize the issue that federal awards have not changed in value since 2003, and financial support for graduate students has remained stagnant over the past 20 years.

4. **Index and evaluate graduate student funding on a regular basis to ensure the sustainability of research infrastructure in Canada**

To ensure the long-term sustainability of research in Canada and appropriately retain and attract top students, we recommend indexing the value of stipends and Tri-Council scholarships to the consumer price index (Statistics Canada 2023). This is already done in countries like Australia, which index all graduate student scholarships annually to adjust for inflation (Khoo 2021). This will better ensure in future that we do not reach a financial crisis among graduate students

5. **Remove limits on work outside of studies for graduate students**

We have highlighted the limitations of working outside of full-time graduate studies. We recommend the removal of certain work limitations (e.g., the 10 h rule), which severely impedes graduate students from obtaining needed supplemental income. To be clear, we *do not* suggest removing these limitations as a way of resolving graduate student financial struggle, as this would merely provide a band-aid solution to a much larger problem that is outlined in this report. Instead, this recommendation should be interpreted as a chance for students with increased financial needs to have the option to seek work outside of their studies while still maintaining their commitment to full-time studies.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Table 3. Recommended stipends for Canadian universities.

University	Total tuition	Rent (one-bedroom) per month	Master's minimum stipend	PhD minimum stipend
University of Alberta	$7086.54	$1020.00	$25 926.54	$31 926.54
University of British Columbia	$5777.02	$2500.00	$42 377.02	$48 377.02
University of Calgary	$7895.83	$1610.00	$33 815.83	$39 815.83
Dalhousie University	$11 060.24	$1800.00	$39 260.24	$45 260.24
Université Laval	$3187.44	$1100.00	$22 987.44	$28 987.44
University of Manitoba	$5011.86	$1090.00	$24 691.86	$30 691.86
McGill University	$6389.50	$1500.00	$30 989.50	$36 989.50
McMaster University	$8389.02	$1640.00	$34 669.02	$40 669.02
Universitié de Montreal	$5207.91	$1500.00	$29 807.91	$35 807.91
University of Ottawa	$7850.76	$1720.00	$35 090.76	$41 090.76
Queen's University	$9110.10	$1700.00	$36 110.10	$42 110.10
University of Saskatchewan	$5873.97	$1020.00	$24 713.97	$30 713.97
University of Toronto	$7997.28	$2300.00	$42 197.28	$48 197.28
University of Waterloo	$8414.33	$1800.00	$36 614.33	$42 614.33
University of Western Ontario	$9088.19	$1680.00	$35 848.19	$41 848.19

Note: Recommended values for minimum master's and PhD stipends at 15 Canadian universities, based on tuition fees, average rent prices for a one-bedroom apartment in corresponding cities, plus an additional $550 per month for expenses and additional $500 for PhD students per month. Additional details on calculations, sources, etc. can be found in Supplementary Table 3.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

6. **Further investigate the role of EDI factors in financial struggle**

One limitation of our survey is that we did not have sufficient demographic data to provide tangible and targeted EDI-centred recommendations. Future work should investigate the financial status of graduate students who identify as members of historically underrepresented groups, including women, Indigenous peoples (First Nations, Inuit, and Métis), persons with disabilities, members of visible minority/racialized groups, and members of LGBTQ2+ communities This will enable a more diverse understanding of students' needs depending on their demographic background and the subsequent development of effective, equitable, and evidence-based programs and(or) policies to support students from historically underrepresented groups.

Acknowledgements

We would like to acknowledge all of the graduate students across Canada who participated in our survey, as well as the various student associations, organisations, and individuals who helped distribute the survey. We would also like to thank Amy Johnston for assistance with analysis of demographic cross-sections, as well as all those from the Science Policy Exchange and Toronto Science Policy Network providing insight during the initial development of the survey.

Article information

History dates

Received: 27 January 2023
Accepted: 10 March 2023
Accepted manuscript online: 11 April 2023
Version of record online: 11 April 2023

Copyright

© 2023 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Data availability

Data generated or analysed during this study are not publicly available due to our stated privacy policy to ensure anonymity of respondent, but anonymized data may be available from the corresponding author on reasonable request.

Author information

Author ORCIDs

Sarah Jane Laframboise https://orcid.org/0000-0003-1731-4391
Mercedes Rose https://orcid.org/0000-0002-2350-8793
Matthew D. Berg https://orcid.org/0000-0002-7924-9241
Sami Aftab Abdul https://orcid.org/0000-0001-8264-435X
Garrett Fairman https://orcid.org/0000-0002-9271-6991
Jacky Deng https://orcid.org/0000-0002-2633-6624
Katherine Shaw https://orcid.org/0000-0002-5440-6079
Alexa D'Addario https://orcid.org/0000-0003-3369-0696

Author notes

Sarah Jane Laframboise, Thomas Bailey, Anh-Thu Dang, Mercedes Rose, and Zier Zhou contributed equally.

Author contributions

Conceptualization: SJL, TB, AD, MR, KO, SE, DMB, GF
Data curation: SJL, TB, AD, MR, ZZ, SAA, KO, SE, DMB, GF
Formal analysis: SJL, TB, AD, MR, ZZ, SAA, JD, KS

Investigation: SJL, TB, AD, MR, ZZ, SAA, KO, SE, DMB, GF, KS
Methodology: SJL, TB, AD, MR, SAA, KO, SE, DMB, GF, JD, KS
Project administration: SJL, TB, AD, MR
Resources: SJL, TB, AD, MR
Supervision: SJL, TB, AD, MR
Validation: SJL, TB, AD, MR, ZZ, MDB, SH, KS
Visualization: SJL, TB, AD, MR, ZZ, SAA
Writing – original draft: SJL, TB, AD, MR, ZZ, MDB, SH, SAA, JD, KS, NN, AD, ME, KS
Writing – review & editing: SJL, TB, AD, MR, ZZ, MDB, SH, SAA, KO, JD, KS, NN, AD, ME, KS

Competing interests

The authors declare there are no competing interests.

Funding information

The authors declare no specific funding for this work.

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/bcb-2023-0021.

References

Baskaran, S., Maddiboina, D., Kum, J.J.Y., Reuben, R., Kharas, K., Bukuroshi, E., et al. 2021. Improving the accessibility of Federal Graduate Research Awards in Canada. J. Sci. Policy Governance, **18**(04). Available at: doi:10.38126/JSPG180405.

BC Gov News. 2021. Graduate scholarship awards supporting student excellence in B.C. Available from https://news.gov.bc.ca/releases/2021AEST0023-000579 [accessed 8 March 2023].

Canadian Institutes of Health Research. 2014. Evaluation of the Vanier Canada Graduate Scholarships Program Final report 2014. Available from https://cihr-irsc.gc.ca/e/documents/vanier-evaluation-report-en.pdf [accessed 7 January 2023].

Chen, C. 2022. Canadian rent report. Available from https://www.zumper.com/blog/rental-price-data-canada/ [accessed 7 January 2023].

Communications Nova Scotia. 2023. Graduate scholarships support research, spur innovation. Available from https://novascotia.ca/news/release/?id=20140425001 [accessed 8 March 2023].

Council of Canadian Academies. 2021. Degrees of success—the expert panel on the labour market transition of PhD graduates. Available from https://cca-reports.ca/wp-content/uploads/2021/01/Degrees-of-Success_FullReport_EN.pdf [accessed 7 January 2023].

Dalhousie Working for Inclusion in Chemical Sciences. 2021. Available from https://twitter.com/Dal_WIC/status/1425411679463038977 [accessed 7 January 2023].

Department of Finance Canada. 2022. Budget 2022. Available from https://budget.canada.ca/2022/report-rapport/toc-tdm-en.html [accessed 7 January 2023].

Employment and Social Development Canada. 2022a. Canada student grant for students with disabilities. Available from https://www.canada.ca/en/employment-social-development/services/education/grants/disabilities.html [accessed 5 January 2023].

Financial Consumer Agency of Canada. 2022. Setting up an emergency fund. Available from https://www.canada.ca/en/financial-consumer-agency/services/savings-investments/setting-up-emergency-funds.html [accessed 5 January 2023].

Gerbing. 2021. Enhancement of the command-line environment for use in the Introductory Statistics Course and beyond. J. Stat. Data Sci. Educ. **29**(3): 251–266. doi:10.1080/26939169.2021.1999871.

Government of Alberta. n.d. Alberta student aid. Available from https://studentaid.alberta.ca/scholarships/alberta-graduate-excellence-scholarship/ [accessed 8 March 2023).

Government of Canada. 2019. Dimensions: equity, diversity and inclusion. Available from http://www.nserc-crsng.gc.ca/NSERC-CRSNG/EDI-EDI/Dimensions-Charter_Dimensions-Charte_eng.asp [accessed 5 January 2023].

Immigration, Refugees and Citizenship Canada. 2022. International students to help address Canada's labour shortage. Available from https://www.canada.ca/en/immigration-refugees-citizenship/news/2022/10/international-students-to-help-address-canadas-labour-shortage.html [accessed 5 January 2023].

Khoo, S. 2021. How Canada short-changes its graduate students and postdocs. Available from https://www.universityaffairs.ca/opinion/in-my-opinion/how-canada-short-changes-its-graduate-students-and-postdocs/ [accessed 24 January 2023].

Lariviere, V. 2012. On the shoulders of students? The contribution of PhD students to the advancement of knowledge. Scientometrics, **90**: 463–481. doi: h10.1007/s11192-011-0495-6.

Likert, R. 1932. A technique for the measurement of attitudes. Arch. Psychol. **22**(140): 5–55.

McGill University. 2022. Guideline on hours of work. Available from https://www.mcgill.ca/study/2022-2023/university_regulations_and_resources/graduate/gps_gi_guideline_hours_of_work#:~:text=In%20order%20to%20maintain%20full,with%2012%20hours%20per%20week. [accessed 8 March 2023].

Marhnouj, S. 2022. Ottawa grad students, postdocs facing 'huge financial burden', CBC News. Available from https://www.cbc.ca/news/canada/ottawa/graduate-students-affordability-living-costs-1.6551913 [accessed 5 January 2023].

National Sciences and Engineering Research Council of Canada. 2022. Guidelines on the assessment of contributions to research, training and mentoring. Available from https://www.nserc-crsng.gc.ca/NSERC-CRSNG/Policies-Politiques/assessment_of_contributions-evaluation_des_contributions_eng.asp#1.1 [accessed 5 January 2023].

Ontario Council on Graduate Studies. 2021. Principles for graduate study at Ontario universities. Available from https://cou.ca/wp-content/uploads/2021/11/OCGS-Principles-for-Graduate-Study-at-Ontarios-Universities-October-2021.pdf [accessed 7 January 2023].

R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Seabold, S., and Perktold, J. 2010. Statsmodels: econometric and statistical modelling with Python. *In* Proceedings of the 9th Python in Science Conference.

Scherer, S., and Smith, F. 2022. Canada annual inflation gallops to near 40-year high of 7.7% in May. Reuters. Available from https://www.reuters.com/business/finance/view-canada-annual-inflation-rate-accelerates-77-may-highest-39-years-2022-06-22/ [accessed 7 January 2023].

Science & Policy Exchange. 2020. Rethinking federal research funding: towards more equitable funding for Canada's next generation. Available from https://www.sp-exchange.ca/rethinking-federal-research-funding [accessed 7 January 2023].

Statistics Canada. 2018. Canadian survey on disability, 2017. Available from https://www150.statcan.gc.ca/n1/en/daily-quotidien/181128/dq181128a-eng.pdf?st=-SoDoz3L [accessed 8 January 2023].

Statistics Canada. 2019. Postsecondary graduates, by province of study and level of study. Table 37-10-0030-01. Available from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710003001&pickMembers%5B0%5D=1.1&cubeTimeFrame.startYear=2000&cubeTimeFrame.endYear=2015&referencePeriods=20000101%2C20150101. [accessed 8 January 2023].

Statistics Canada. 2020. Survey of postsecondary faculty and researchers, 2019. Available from https://www150.statcan.gc.ca/n1/daily-quotidien/200922/dq200922a-eng.htm. [accessed 8 January 2023].

Statistics Canada. 2022a. Indigenous population continues to grow and is much younger than the non-Indigenous population, although the pace of growth has slowed. Available from https://www150.statcan.gc.ca/n1/en/daily-quotidien/220921/dq220921a-eng.pdf?st=VzkBQ_Y0 [accessed 8 January 2023].

Statistics Canada. 2022b. Market basket measure (MBM) thresholds for the reference family by market basket measure region, component and base year. Table 11-10-0066-01. Available

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

from https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/tab/index-eng.cfm?ID=T2_2 [accessed 8 January 2023].
Statistics Canada. 2022c. Proportion of Canadian and international student enrolments, by International Standard Classification of Education. Table 37-10-0163-03. Available from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710016303 [accessed 8 January 2023].
Statistics Canada. 2022d. To buy or to rent: the housing market continues to be reshaped by several factors as Canadians search for an affordable place to call home. Available from https://www150.statcan.gc.ca/n1/daily-quotidien/220921/dq220921b-eng.htm [accessed 8 January 2023].
Statistics Canada. 2022e. Tuition and living accommodation costs survey. Available from https://www150.statcan.gc.ca/n1/daily-quotidien/220907/dq220907b-eng.htm [accessed 8 January 2023].
Statistics Canada. 2022f. Postsecondary student information system (PSIS). Available from https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2022019-eng.htm [accessed 24 January 2023].
Statistics Canada. 2023. Consumer price index portal, Statistics Canada. Available from https://www.statcan.gc.ca/en/subjects-start/prices_and_price_indexes/consumer_price_indexes [accessed 24 January 2023].
Support our Science. 2022. Written Submission for the Pre-Budget Consultations in Advance of the 2023 Budget. Available from https://www.supportourscience.ca/post/pre-budget-consultations-for-budget-2023-briefing-by-support-our-science [accessed 8 January 2023].
Toronto Science Policy Network. 2020. The early impacts of COVID-19 on graduate students across Canada. Available from https://tspn.ca/covid19-report/. [accessed 8 January 2023].
University of Waterloo. 2022. Guidelines for graduate student employment. Available from https://uwaterloo.ca/graduate-studies-postdoctoral-affairs/current-students/guidelines-graduate-student-employment [accessed 5 January 2023].
Wall, K. 2020. COVID-19 pandemic: financial impacts on postsecondary students in Canada. Statistics Canada. Available from https://www150.statcan.gc.ca/n1/en/pub/45-28-0001/2020001/article/00016-eng.pdf?st=cY63aU8V [accessed 8 January 2023].
Woolston, C. 2022. 'Not even enough money for food': graduate students face cash crunch. Nature, **611**: 189–191. doi: h10.1038/d41586-022-03478-x [accessed 8 January 2023]. PMID: 36316470.

Biochem. Cell Biol. Downloaded from cdnsciencepub.com by 174.115.154.151 on 04/18/24

Chapter 11: Prioritizing Mentorship as Scientific Leaders

11.1 Introduction to the research article presented in this chapter

In 2022, I was selected from thousands of applicants from around the world to be a member of the 2022 Chemical Abstract Services (CAS) Future Leaders. The program brings together early-career researchers in the chemical sciences from around the world, supporting them with workshops and seminars on scientific leadership, communication, and more.

Chapter 11 describes an editorial I led in collaboration with members of the 2022 CAS Future Leaders, published in ACS Central Science. The goals of the article were to communicate to readers: (1) the importance and impacts of mentorship in chemistry; (2) strategies and approaches for engaging in effective and inclusive mentorship in chemistry; and (3) strategies and approaches for creating a chemistry community and culture that prioritizes effective and inclusive mentorship.

Thank you to my co-authors on this manuscript—future (and present!) scientific leaders from across the globe—for their patience and support throughout the two years of drafts and revisions. The team was composed of researchers I deeply admire and respect, and it gives me so much hope for our community that these people care about mentorship in chemistry/science as much as I do. I hope that the piece serves as valuable resource for all mentors and mentees in science and chemistry.

11.2 Copyright

Reprinted from "Deng, J. M.; Ahmed, S. E.; Awoonor-Williams, E.; Banerjee, P.; Barecka, M. H.; Bickerton, L. E.; Di Pietro, S. A.; Dorn, S. K.; Jablonka, K. M.; Laudadio, G.; Kreidt, E.; Mannochio-Russo, H.; Terra, J.; Wilkins, O. H.; Yerneni, S. S.; Yusuf, M. Prioritizing Mentorship as Scientific Leaders. *ACS Cent. Sci.* **2024***, 10* (2), 209–213

doi.org/10.1021/acscentsci.3c00500" with permission from the American Chemical Society.

Copyright © 2024 American Chemical Society.

ACS central science

Open Access
This article is licensed under CC-BY 4.0

EDITORIAL

Prioritizing Mentorship as Scientific Leaders

Cite This: ACS Cent. Sci. 2024, 10, 209−213

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

Downloaded via 174.115.154.151 on April 18, 2024 at 21:32:56 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Sustaining innovation in science, technology, engineering, mathematics, and medicine (STEMM) requires collaborative problem-solving, particularly as we work to address complex global issues.[1] Effective mentoring strategies are essential in accomplishing this goal by helping support diverse talent that will strengthen the future of STEMM.[2] Multiple studies have shown that mentorship is positively related to graduate students' persistence in research productivity, research self-efficacy, rate of degree completion, and program satisfaction.[3−6]

Despite its importance, mentorship is often an overlooked and underappreciated component of scientific training. In this article, we—a diverse authorship team representing ten countries and various chemistry subdisciplines brought together by the CAS Future Leaders program—share mentorship experiences and actions that readers can use to promote more inclusive and productive mentor-mentee relationships in chemistry. We share our insights by reflecting on three key questions:

(1) Why is mentorship important?

(2) What have been some impactful mentorship strategies we have used or experienced?

(3) How can we create a chemistry community that values and prioritizes effective mentorship?

The following sections share patterns across our diverse experiences and perspectives. Full responses from the authors can be found in the Supporting Information.

1. WHY IS MENTORSHIP IMPORTANT?

Reflecting on our collective experiences, our responses to this question identified the following themes. (1) Mentors play a vital role in equipping mentees with key *scientific* skills, such as experimental techniques or how to think critically about their research and careers. (2) Mentors are also crucial in helping mentees develop *professional* skills, such as networking and navigating careers. (3) Lastly, mentors play a critical role in supporting mentees in navigating research challenges and developing resilience (Figure 1).

1.1. Mentors Teach Us How to Be Scientists. *1.1.1. Building Our Scientific Skills.* Scientific mentors play a vital role in imparting essential tacit knowledge concerning research methodologies, experimental design, and data analysis. Through their guidance, mentees can cultivate a solid foundation for conducting research.

1.1.2. Fostering Our Scientific Thinking. Mentors not only facilitate the acquisition of scientific skills but also cultivate a self-belief in mentees that they are scientists. They inspire mentees to embrace critical thinking, pose questions, and challenge assumptions. This process equips mentees with the capacity to assess existing knowledge, pinpoint gaps, and propose innovative experimental approaches. Furthermore, effective mentors can contribute to the development of problem-solving proficiency by sharing their personal experiences in overcoming research hurdles.

1.2. Mentors Teach Us More than Science. *1.2.1. Supporting Our Personal and Professional Development.* Scientific mentors extend their guidance beyond scientific knowledge, offering valuable insights into "soft" skills, such as effective communication, presentation techniques, and scientific writing. They can impart valuable life skills, such as time management, goal setting, or modeling a growth mindset when challenges arise.

1.2.2. Helping Us Navigate Career Paths. Mentors play a pivotal role in helping mentees navigate the range of career choices, both within academia and beyond. Their experience can significantly help with decisions related to pursuing advanced degrees, transitioning into industry roles, or engaging in science policy.[7] Mentors can also help mentees discover diverse career opportunities by facilitating introductions to their professional network, collaboration opportunities, and exposure to different research environments and perspectives.

1.3. Mentors Can Influence Who Continues in Science. *1.3.1. Inspiring and Motivating Mentees.* Science and research rarely follow a straightforward, linear path. Mentors serve as role models for their mentees, and their approach and attitude toward science, as well as their ability to support and encourage mentees, can motivate individuals to pursue and persist in science.

1.3.2. Supporting Scientists from Historically Underserved Groups. Mentors play a vital role in supporting individuals from historically underserved groups in science. They can offer the essential support, understanding, and guidance that are needed to overcome barriers and biases in the scientific community (e.g., the hidden curriculum, implicit expectations). Their role extends

Published: January 10, 2024



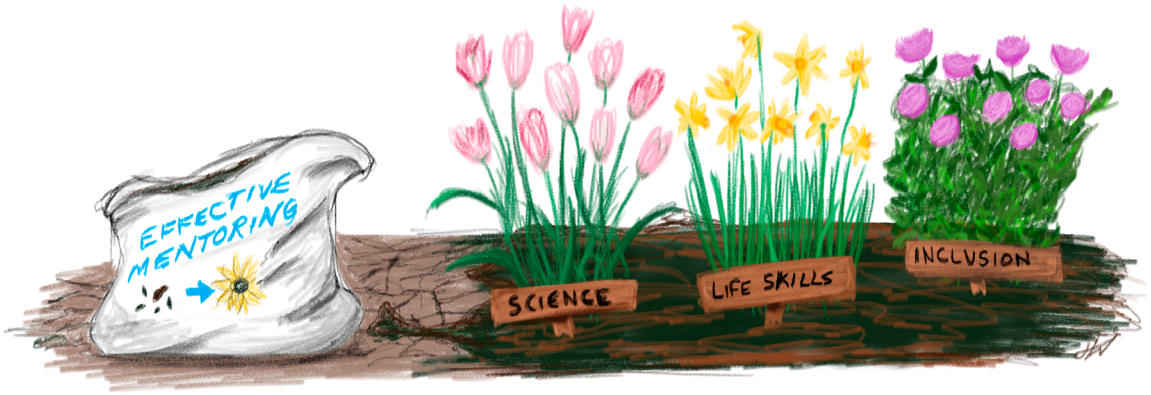


Figure 1. Effective mentorship fosters not only growth in science but professional and personal skills and inclusion. Figure created by Olivia Harper Wilkins.

to advocating for systemic changes that work to disrupt and dismantle inequities faced by underserved groups, contributing to a more inclusive scientific community.

2. WHAT HAVE BEEN SOME IMPACTFUL MENTORSHIP STRATEGIES WE HAVE USED OR EXPERIENCED?

Every mentoring relationship is unique. However, from our experiences, certain strategies can serve as a foundation for fostering effective mentor-mentee interactions. These include (1) setting clear expectations, (2) establishing rapport, (3) maintaining open communication, and (4) empowering mentees.[3]

2.1. Aligning Expectations. *2.1.1. Defining Roles and Responsibilities.* We all found that it was important to have a conversation at the start of our mentorship relationships to outline what each person's roles and responsibilities were. Our most effective mentors have always clarified early on how often we would meet, what kind of feedback they would provide, and what we, as mentees, were expected to contribute.

2.1.2. Setting Clear Timelines and Goals. Establish specific, measurable, achievable, relevant, and time-bound (SMART) goals, such as setting specific timelines for completing a literature review or reaching a specific phase of a project. For example, when handling lab projects, we worked with our mentors to set a goal for completing the initial experiments within an agreed-upon time period. Having these time-bound goals keeps everyone on track.

2.1.3. Discussing Short- and Long-Term Goals. Plan sessions to talk about both immediate and future goals. For example, one could discuss the short-term goal of completing a literature review by the end of the month and the long-term goal of presenting findings at a conference in six months. This way, everyone is aware of and working toward the same objectives.

2.1.4. Track Progress Together. Keep a shared document or tool to track progress collaboratively. This could be a simple spreadsheet or document where you note what tasks have been completed, what's in progress, and what's coming up. This real-time tracking has helped us and our mentors stay informed about our progress.

2.2. Building Rapport. *2.2.1. Choosing Kindness.* Regardless of any other label we may give ourselves, we are human beings first. Our most effective mentors have always prioritized empathy and kindness. If you're comfortable, invite opportunities for conversations about personal experiences and interests (e.g., hobbies, interests, values). This personal touch cultivates trust and enriches the mentor-mentee relationship. Attending professional (e.g., seminars, workshops) and group events together can also help strengthen such bonds.

2.2.2. Sharing Failures. In research, we place so much emphasis on people being "experts" in something that sometimes we forget that we can't be experts in all areas. Our best mentors have reminded us that we are not expected to know everything (or perhaps even anything). This has helped us feel more secure in our pursuit of knowledge, empowering us to spend our energy stumbling through research rather than pretending we are going through the process flawlessly. It helps when mentors open up about their own past challenges and failures in their research journey, for example, discussing a project that didn't go as planned or an experiment that didn't yield the expected results. This helps mentees understand that setbacks are normal and encourages them to approach research with a growth mindset.

2.2.3. Encouraging Questions. Creating an environment where asking questions is encouraged and that curiosity drives scientific progress. As mentees, some of us may be hesitant to ask a question. Effective mentors recognize this and actively invite mentees to share what they're thinking. This has helped us move past viewing questions as a sign of weakness but rather as an opportunity for growth.

2.3. Maintaining Open Communication. *2.3.1. Scheduling Regular Check-ins.* A brief, regular, agreed-upon check-in period to discuss progress and address any challenges can help maintain open communication between the mentor and the mentee. Consistent communication also helps mentees feel supported.

2.3.2. Listening Actively. Practice active listening by showing genuine interest in mentees' thoughts and concerns. This demonstrates empathy and contributes to building a strong rapport.

2.3.3. Promoting a Culture of Constructive Feedback. Mentorship is a two-way street. Our mentees have given us opportunities to provide feedback on the mentorship experience, helping us establish a two-way communication channel and culture of growth. For instance, they have asked us for our perspective on what aspects of the experience have been helpful and what could be improved. Surveys could also be used to mitigate mentor−mentee power dynamics, allowing mentees to provide their feedback anonymously.

2.4. Facilitating Mentee Agency. *2.4.1. Supporting Self-Reflection.* Our mentors have assigned us self-reflection exercises where, rather than having them evaluate us, we evaluate our own progress, strengths, and goals. These can take

https://doi.org/10.1021/acscentsci.3c00500
ACS Cent. Sci. 2024, 10, 209−213

the form of regular progress reports/updates in which mentors prompt their mentees to reflect on their professional and personal development after a certain period of time, followed by a 1-on-1 discussion.

2.4.2. Encouraging Autonomy and Independence. Our most effective mentors recognized when we were ready for the opportunity to lead projects or experiments. Allowing us to make decisions and take ownership fostered a sense of agency and boosted our confidence. For example, mentors can give mentees independent opportunities to explore the literature, tools, and potential collaborators. If one is mentoring a student on a lab project, invite them to take the lead on planning and executing certain experiments.

3. HOW CAN WE CREATE A CHEMISTRY COMMUNITY THAT VALUES AND PRIORITIZES EFFECTIVE MENTORSHIP?

Although mentorship is crucial for retaining talent and promoting progress in chemistry, it often receives less attention compared to other areas of our profession and culture.[8] Cultural change and community buy-in are needed, which can be accomplished through a systemic approach to incorporating and incentivizing effective mentorship in all aspects of our scientific community and practice.[9,10] We emphasize the importance of highlighting and valuing mentorship through mentorship awards and including mentorship as criterion in applications for scientific scholarships, awards, grants, and tenure applications. There must also be opportunities for scientists to regularly discuss, practice, and reflect on mentorship alongside colleagues and peers. Lastly, broader policy changes are needed to ensure that how our community views and values mentorship is sustained and formalized.

3.1. Highlighting Effective Mentorship. *3.1.1. Recognizing Outstanding Mentors.* Establishing awards or recognition programs that celebrate mentors who have made significant positive impacts on their mentees' careers and growth. For instance, a departmental award could include a category for exceptional mentorship. Additionally, consider integrating mentorship evaluation into existing milestones such as research awards or tenure assessments. Regularly feature mentor profiles, success stories, and interviews in community newsletters, journals, or social media platforms. These mechanisms and initiatives emphasize the importance of mentorship within our chemistry community.

3.1.2. Evaluating Mentorship As a Criterion in Award, Grant, and Job Applications. Explicitly including mentorship as a criterion in award, grant, job, and other applications communicates to members of our chemistry community that mentorship is something we prioritize and view as an *essential* part of being successful scientists. This may also help our community better recognize the contributions members of historically underrepresented groups in science, who have historically and disproportionately been called upon to mentor and support students and trainees.[11] Ways to do this include asking for mentorship statements (perhaps included as part of teaching statements) or inviting reference letters from mentees.

3.1.3. Hosting Panel Discussions and Webinars. Organizing panel discussions and webinars that focus on successful mentorship stories and strategies. These events can showcase the impact of mentorship and provide insights into effective practices. For instance, a panel could discuss how mentorship aided a mentee's transition from academia to industry and the strategies that contributed to that success.

3.2. Creating Opportunities to Learn and Practice Mentorship. *3.2.1. Creating Mentorship Training Workshops.* Host hands-on workshops that provide mentors with practical training on effective mentorship techniques, such as the skills described in this article. These could also include role-playing scenarios where mentors and mentees can practice setting expectations/goals, having challenging conversations, or providing feedback. One example of this is the ACS Mentor Training Workshop for Graduate Students and Postdoctoral Scholars, which helps early career researchers "learn strategies for developing or improving mentoring skills to cultivate productive mentor-mentee relationships".[12] Other examples include the ACS's Postdoc to Faculty and New Faculty Workshops and the RSC's Career Talks.[13−15] Additional resources and workshops are available in the Supporting Information (Laura; pp 20−21).

3.2.2. Establishing Peer Mentorship Circles. Establish small, supportive peer mentorship circles. These circles serve as spaces where mentors and potential mentors come together to discuss hurdles, share personal experiences, and collectively work on their mentorship skills. Through candid conversations, mentors can learn from each other's insights and adapt their strategies accordingly. For instance, a circle could meet monthly to discuss common challenges, compare strategies, and brainstorm effective solutions.

3.2.3. Fostering Diverse Mentor-Mentee Matchmaking. Mentor-mentee matching programs within the chemistry community can help create structured opportunities for mentors and mentees to connect. One example of this is the ACS Career Consultants program, which allows individuals to connect 1-on-1 with experienced ACS members who are willing to share their expertise in specific job markets (academia/industry), processes (such as transitioning into an independent research position or successful retirement), and their passion in seeing the long-term growth and success of their mentees.[16] Other examples include departmental mentorship programs and the Chemistry Women Mentoring Network (CWMN).[17,18]

3.3. Advocating for Policy Changes That Prioritize Mentorship. In order for effective mentorship to be sustained and practiced, it must also be formalized through policy changes. Many of the strategies discussed in this article are aligned with developing individual development plans (IDPs): a systematic approach for helping individuals in their professional growth and career advancement. Required, annual IDPs are becoming increasingly common in graduate programs,[19,20] and their use is reviewed in grant applications for the NIH.[21] These types of policy changes that ensure mentors and mentees have regular conversations around expectations, progress, goals, and career development could help our community begin to view mentorship as an essential part of our profession. ACS and AAAS/*Science* have also developed online tools to support graduate students and postdoctoral researchers in developing IDPs.[22,23]

3.3.1. A Call to Action: Prioritizing Mentorship As Scientific Leaders. Mentorship is an essential aspect of success in science, yet it is often overlooked. To establish a successful mentorship relationship, both mentors and mentees must set clear goals and expectations, build rapport, and facilitate open communication.

How can we make mentorship a more normalized topic in academia and beyond? By engaging in discussions about the advantages of mentorship, we can emphasize its importance as a skill to be developed. And it's not just mentees who benefit—mentors can also gain valuable insights and experiences through the process.

https://doi.org/10.1021/acscentsci.3c00500
ACS Cent. Sci. 2024, 10, 209−213

To encourage and facilitate mentorship, organizations should establish programs and awards that incentivize and recognize effective mentoring. By making mentorship a priority and investing in functional mentoring alliances, we can create a culture of growth and development that benefits everyone involved.

As early career researchers in the chemical sciences, we believe that effective mentorship is foundational for sustained innovation in STEMM fields (Figure 2). We hope to inspire readers to prioritize mentorship in their scientific careers and to recognize its importance through tangible actions. Together, we can create a culture of inclusivity and support that strengthens the future of chemistry.

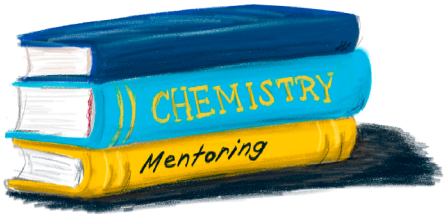


Figure 2. We believe mentoring is an essential component of STEMM education, and we hope that mentorship programs will become a regular part of professional training. Figure created by Olivia Harper Wilkins.

Jacky M. Deng, Department of Chemistry and Biomolecular Sciences, University of Ottawa orcid.org/0000-0002-2633-6624

Salma Elgaili Ahmed, Centre for Advanced Imaging and The Australian Institute for Bioengineering and Nanotechnology, The University of Queensland

Ernest Awoonor-Williams, Novartis Institutes for BioMedical Research orcid.org/0000-0002-9127-8539

Progna Banerjee, Center for Nanoscale Materials, Argonne National Laboratory orcid.org/0000-0003-3257-7317

Magda H. Barecka, Department of Chemical Engineering, Northeastern University College of Engineering; Department of Chemistry and Chemical Biology, Northeastern University College of Science orcid.org/0000-0002-0772-164X

Laura E. Bickerton, EaStCHEM School of Chemistry, University of Edinburgh

Silvina A. Di Pietro, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory

Stanna K. Dorn, Division of Chemistry and Chemical Engineering, California Institute of Technology orcid.org/0000-0002-8644-9962

Kevin Maik Jablonka, Laboratory of Organic and Macromolecular Chemistry (IOMC) and Center for Energy and Environmental Chemistry Jena (CEEC Jena), Friedrich Schiller University Jena

Gabriele Laudadio, Department of Chemistry, Scripps Research orcid.org/0000-0002-2749-8393

Elisabeth Kreidt, Department of Chemistry and Chemical Biology, TU Dortmund University

Helena Mannochio-Russo, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego

Júlio Terra, Laboratory of Sustainable and Catalytic Processing, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne

Olivia Harper Wilkins, NASA Postdoctoral Program Fellow, NASA Goddard Space Flight Center

Saigopalakrishna S. Yerneni, Department of Chemical Engineering, Carnegie Mellon University

Maha Yusuf, Department of Chemical Engineering, Stanford University orcid.org/0000-0001-7908-2915

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.3c00500.

Full responses to mentorship questions from authors and author biographies (PDF)

Author Information

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.3c00500

Author Contributions

The coauthors (the second author to the last author listed) contributed equally to conceptualization and writing (reviewing/editing). O.H.W. – visualization, J.M.D. – writing (original draft), reviewing/editing, supervision.

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

ACKNOWLEDGMENTS

We thank CAS for an incredible opportunity and experience in the CAS Future Leaders 2022 program and for instilling in us a collective drive to be better mentors and leaders. The work performed by author S.A.D. was under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, IM release LLNL-JRNL-848416. The work performed by author S.E.A. was under the auspices of the Australian Government Research Training Program and The University of Queensland.

ABBREVIATIONS

STEMM, science, technology, engineering, mathematics, medicine; RSC, Royal Society of Chemistry; AAAS, American Association for the Advancement of Science; NIH, National Institutes of Health

REFERENCES

(1) United Nations. Transforming our World: the 2030 Agenda for Sustainable Development, https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

(2) Hofstra, B.; Kulkarni, V. V.; Galvez, S. M. N.; He, B.; Jurafsky, D.; McFarland, D. A. The Diversity–Innovation Paradox in Science. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (17), 9284–9291.

(3) Paglis, L. L.; Green, S. G.; Bauer, T. N. Does Adviser Mentoring Add Value? A Longitudinal Study of Mentoring and Doctoral Student Outcomes. *Res. High. Educ.* **2006**, *47* (4), 451–476.

(4) Lunsford, L. Doctoral Advising or Mentoring? Effects on Student Outcomes. *Mentor. Tutoring Partnersh. Learn.* **2012**, *20* (2), 251–270.

(5) Williams, S. W.; Thakore, B. K.; McGee, R. Coaching to Augment Mentoring to Achieve Faculty Diversity: A Randomized Controlled Trial. *Acad. Med.* **2016**, *91* (8), 1128–1135.

(6) Mcgee, R.; Keller, J. L. Identifying Future Scientists: Predicting Persistence into Research Training. *CBE Life Sci. Educ.* **2007**, *6*, 316–331.

(7) John, T.; Cordova, K. E.; Jackson, C. T.; Hernández-Mondragón, A. C.; Davids, B. L.; Raheja, L.; Milić, J. V.; Borges, J. Engaging Early-Career Scientists in Global Policy-Making. *Angew. Chem.* **2023**, *135*. e202217841.

(8) National Academies of Science Engineering and Medicine. *The Science of Effective Mentorship in STEMM*; Byars-Winston, A., Dahlberg, M. L., Eds.; National Academies Press: Washington, D.C., 2019.
(9) Reinholz, D. L.; Apkarian, N. Four Frames for Systemic Change in STEM Departments. *Int. J. STEM Educ.* **2018**, *5* (1). DOI: 10.1186/s40594-018-0103-x.
(10) Stachl, C. N.; Brauer, D. D.; Mizuno, H.; Gleason, J. M.; Francis, M. B.; Baranger, A. M. Improving the Academic Climate of an R1 STEM Department: Quantified Positive Shifts in Perception. *ACS Omega* **2021**, *6* (22), 14410−14419.
(11) Griffin, K. A. Institutional Barriers, Strategies, and Benefits to Increasing the Representation of Women and Men of Color in the Professoriate. In *Higher Education: Handbook ofTheory and Research*; Perma, L. W., Ed.; Springer Nature Switzerland: Switzerland, 2020; p 331. DOI: 10.1007/978-3-030-31365-4_4.
(12) American Chemical Society. ACS Mentor Training Workshop for Graduate Students and Postdoctoral Scholars, https://www.acs.org/education/students/graduate/mentor-training.html (accessed Aug 10, 2023).
(13) American Chemical Society. Postdoc to Faculty Workshop, https://www.acs.org/education/students/graduate/gettingready/academiccareers/postoral-workshop-for-prospective-chemistry-faculty.html (accessed Oct 23, 2023).
(14) American Chemical Society. New Faculty Workshop, https://www.acs.org/education/educators/coursesworkshops/higher-ed-workshops/new-faculty-workshop.html. The New Faculty Workshops focus, learning strategies in the classroom. (accessed Oct 23, 2023).
(15) Royal Society of Chemistry. Career Talks https://www.rsc.org/careers/career-talks-events/#supporting-research (accessed Oct 23, 2023).
(16) American Chemical Society. ACS Career Consultants.
(17) Chemistry Women Mentoring Network. Chemistry Women Mentoring Network https://brandicossairt.wixsite.com/chemwmn (accessed Oct 25, 2023).
(18) uOttawa Chemistry and Biomolecular Graduate Students' Association. Buddy Program.
(19) Utah State University College of Science. Graduate Student Individual Development Plans https://www.usu.edu/science/students/graduate/individual-development-plan (accessed Oct 23, 2023).
(20) UC Irvine Department of Chemistry. Individual Development Plan https://www.chem.uci.edu/graduate/current-students/IDP (accessed Oct 23, 2023).
(21) National Institute of Health. Individual Development Plans https://www.nigms.nih.gov/training/strategicplanimplementationblueprint/Pages/IndividualDevelopmentPlans.aspx (accessed Oct 23, 2023).
(22) American Chemical Society. ChemIDP https://chemidp.acs.org/ (accessed Oct 23, 2023).
(23) Science Careers. myIDP https://myidp.sciencecareers.org/ (accessed Oct 23, 2023).

https://doi.org/10.1021/acscentsci.3c00500
ACS Cent. Sci. 2024, 10, 209−213

Chapter 12: Conclusions and Final Reflection

This chapter provides an overview of my doctoral research projects, including their findings, implications, challenges, and future directions. I conclude with a final reflection on my doctoral studies.

12.1 Overview of research significance and implications

We live in a world facing a myriad of complex global challenges that will require all citizens to be able to make and justify decisions using scientific evidence (Social Sciences and Humanities Research Council, 2018; United Nations, 2015). We also live in an increasingly global society, one in which linguistic diversity in chemistry classrooms and research spaces is prevalent and increasing (Crossman et al., 2022; Higher Education Strategy Associates, 2020; Institute of International Education, 2020; National Academies of Science Engineering and Medicine, 2018). If we are to succeed in solving the challenges facing our world, we must learn how to best equip students with the scientific skills required to navigate these challenges, while also ensuring that all students have equal opportunities to learn, regardless of their language background. The absence of research on these topics would impede progress in scientific literacy and inclusion, limiting our capacity to navigate and solve pressing global issues. Using qualitative, quantitative, and mixed research methods, my doctoral work has contributed to addressing two overarching questions in chemistry education research: (1) How do students engage in argumentation and reasoning in chemistry education? and (2) How do students from diverse language backgrounds learn, communicate, and do chemistry?

My research has uncovered that how students' construct arguments about chemical phenomena—including the reasoning, granularity, and comparisons in arguments—can depend on different curricular factors and expectations (Chapters 2 and 3) (Bodé et al., 2019; Deng & Flynn, 2021). This research was amongst the first to determine (1) How students construct

arguments when justifying phenomena in organic chemistry and (2) How different curricular factors and expectations, such as the prompts and course learning outcomes, may shape students' approaches to argumentation. We found that how students argue is dynamic and shaped not only by their chemistry knowledge and motivation, but also by factors related to course structure and expectations. Furthermore, much of the previous research in chemistry education had also solely focused on students' reasoning in their arguments (Becker et al., 2016; Cooper et al., 2016; Moon et al., 2016, 2019; Osborne & Patterson, 2011; Russ et al., 2008; Watts et al., 2020; Weinrich & Talanquer, 2016). Our work was the first to apply a framework for analyzing students' arguments that included other dimensions, such as levels of granularity and levels of comparison. From this work, we also shared ways researchers and educators can use constructive alignment and backwards design to support students in developing and demonstrating their reasoning and argumentation skills across a curriculum (Chapter 4) (Deng, Carle, et al., 2022).

Our work on linguistic equity in chemistry education has been foundational in understanding the learning and research experiences of chemistry students and trainees from diverse language backgrounds. This work was guided by two key insights. First, we recognized that chemistry classrooms were linguistically diverse, and will likely become more diverse as higher education becomes more globalized and international (Crossman et al., 2022; Higher Education Strategy Associates, 2020; Institute of International Education, 2020; National Academies of Science Engineering and Medicine, 2018). Second, we recognized that constructing a scientific argument can be a linguistically complex task, requiring one to not only *know* how to justify, but to *communicate* that justification in the English language (Lee & Fradd, 1998; McNeill et al., 2006; McNeill & Berland, 2010; McNeill & Krajcik, 2008a, 2008b). Despite this tension, no other research in either chemistry or science education research had, at the time, had investigated the role that students' language backgrounds might play in how they construct arguments. Our research in Chapter 5 found that English proficiency and history is associated with how students justify chemical phenomena. Follow-up interviews with fifteen participants revealed that

students' comfort communicating in English on assessments is complex and may affected by a combination of general and academic language experiences. Our findings suggest a need to identify the barriers encountered by students from diverse language backgrounds in chemistry classrooms to ensure all students equitable ways to demonstrate their abilities (Chapter 5) (Deng, Rahmani, et al., 2022)..

Building on what we found in the research described in Chapter 5, the latter parts of my doctoral work shifted to a broader focus on the role of language in in chemistry learning and research (Chapters 6 and 7). In focus groups and interviews with 18 Eng+ chemistry trainees (Chapter 6), we found that chemistry trainees learning English-as-an-additional language (Eng+) face pervasive linguistic challenges and barriers in research. These trainees also demonstrated resilience and a growth mindset and shared implications for how supervisors and institutions can better support them. Our preliminary findings from a questionnaire study with 219 chemists from a different language backgrounds and career stages suggests that similar challenges and desire for more support may be prevalent in the broader chemistry community (Chapter 7). In addition to shining a light on the experiences of researchers and learners from diverse language backgrounds, my work in these chapters has also contributed to the relatively under-studied area of chemistry graduate student education—an area of study essential for advancing and sustaining innovation in chemistry (Jaquette et al., 2016; Posselt & Grodsky, 2017). The findings from these projects present broad implications for approaches, strategies, and resources supervisors and institutions can enact that may improve linguistic equity in the chemical sciences.

Collectively, the work described in my thesis has contributed to many areas of chemistry education research, including supporting students reasoning in chemistry, linguistic equity in chemistry, as well as instrument development in chemistry education research, and chemistry graduate student education. Collectively, the projects I have been a part of and led during my doctoral studies have contributed to many different areas of chemistry education research and

beyond. At the time of writing, these manuscripts, with research spanning over five years, have been cited over 160 times and viewed over 25,000 times. Our work on reasoning and argumentation in chemistry education has informed many other studies and initiatives, including investigations of students' reasoning in other chemistry contexts and the development of new teaching approaches and formative assessment tools to support the development of students' reasoning (*e.g.,* scaffolding, machine learning) (Dood & Watts, 2022; Lieber et al., 2022; Martin et al., 2023; Martin & Graulich, 2024). Our work on language in chemistry education and research is amongst the first in the field, and it will serve as a basis for many future initiatives and studies on linguistic equity in chemistry and science. Our methods manuscript on response process validity evidence in CER has already been a valuable resource for the CER community, and it will continue to be a foundational text as our field continues to grow. The papers on chemistry graduate student EDI communities, financial challenges of graduate students in Canada, and mentorship in chemistry have shaped and will continue to shape conversations about the importance of graduate student leadership and education in science.

All research comes with limitations and challenges. Rather than repeat the limitations of each project, which are described in detail in each chapter, I will share here my thoughts on what I wish I had done more of or differently across my research projects as a graduate student in CER.

I believe that what makes CER study valuable is not only whether it inspires and informs future research, but whether it has a tangible impact on teaching and learning in chemistry. Though my thesis projects have been significant contributions to the CER literature, I have not conducted any research, nor do I have any data on whether what I have done as a researcher has had a tangible impact on teaching and learning in chemistry. Though I have engaged in knowledge mobilization activities—I have given talks, hosted workshops and seminars, and written pieces for different types of audiences and goals related to my work—any claims I might have about the whether these activities have tangibly shaped perspectives and practices would be presumptuous. Therefore, one overarching limitation of my work is that we do not know how much of an impact

(if any) our work has had on chemistry students and educators—in other words, those who stand to benefit the most from our research. If I had more time or could repeat my studies, I would likely spend more time on knowledge mobilization, working with educators and students to better understand how our research findings can be translated into their day-to-day teaching and learning, and then evaluating the impacts of new approaches or initiatives. This challenge is not unique to my research—I detail the challenges of what is called the "research–practitioner gap" in the following section and reference here recent work in our field that suggests a growing awareness of and need to address these types of gaps.

12.2 Overview of future research directions

My view of chemistry education research has constantly evolved during my doctoral studies. In my conversations with other chemistry education researchers, as well as through my observations of recent trends in the CER literature, I believe that the field of CER is experiencing a significant shift in its foci. When I first started as a graduate student, much of CER was focused on students' cognitive understanding of chemistry concepts and knowledge (see inherent CER, Taber 2013). The literature at the time was filled with research on students' alternative conceptions, concept inventories, and learning strategies to support the development of students' chemistry content knowledge.

With time, new research and researchers, and major events such as the 2020 anti-racism movement, CER has become more focused on equity, diversity, and inclusion in chemistry education (Ryu et al., 2021; Wilson-Kennedy et al., 2020; Winfield et al., 2020). Rather than thinking about what happens within the individual inside the chemistry classroom, our field appears to be becoming increasingly interested in the systemic forces that shape learning inside and outside the chemistry classroom. I believe that this shift has, overall, been beneficial for the field and for broadening the impact of the research that we do. I believe that the shift represents a collective understanding that students are not simply isolated vessels for us as educators to impart

chemistry knowledge into, but rather dynamic and diverse individuals whose learning experiences in chemistry are shaped by a multitude of factors both inside and outside the classroom. My work has followed this shift: I began my studies focused on how student reasoning and argue in chemistry, and then gradually began to focus more on how students' language backgrounds (*i.e.*, a part of students' identities and lives beyond the chemistry classroom) shaped their learning and research experiences. My work has also been among the first to explore EDI within the context of chemistry graduate education—an increasing prominent but historically understudied area of chemistry education research.

As our field becomes more EDI-focused, I believe that our work on linguistic equity in chemistry education will serve as a valuable foundational resource for researchers, especially as the broader chemistry community becomes more culturally and linguistically diverse. Future work could evaluate the efficacy of interventions to support learners and trainees from diverse language backgrounds, such as those proposed by participants in our research (e.g., chemistry writing centres, social networking opportunities). I also believe that technology will also play a larger role in our future research as a field. The emergence of artificial intelligence and machine learning tools has already begun to spark new conversations and work in CER and education more broadly (Alasadi & Baiz, 2023; Emenike & Emenike, 2023), including our own work described in Chapter 7 and new research on student reasoning and linguistic equity in chemistry (Martin et al., 2023; Martin & Graulich, 2024).. As these tools continue to evolve, so too will our relationship to them as chemistry students, educators, and education researchers. Additionally, I suggest caution in viewing these tools as a panacea for "solving" the challenges faced by those from diverse language backgrounds. Regardless of what these tools may offer now and in the future, our research suggests there will likely always be a need for empathetic and effective mentors and colleagues in chemistry research and training.

The final comment I have about future research directions for CER is one that I think is relevant to all research we do in the field moving forward. During my seven years in CER as an undergraduate and graduate student, I have noticed that we, unlike any other subfield of chemistry, must constantly justify our existence and right to be members of the broader chemistry community. The sense I have gotten is that we are largely viewed as social scientists masquerading as chemists, doing research that can be opaque and foreign to many of our colleagues. Though I believe that CER is as essential as any other chemistry subdiscipline in supporting innovation and growth in science, I can also empathize with those who question whether what we do is "really chemistry". In CER, we often present our research findings and implications in ways that frame educators as autonomous entities with infinite resources, time, and flexibility, when reality often looks like opposite. In other words, there is a gap between what we do and find as researchers and what is feasible for practitioners (*i.e.*, the research–practitioner gap) (Herrington et al., 2019; Szteinberg et al., 2014). I believe that the future growth of CER, including our field's impact on the lives and experiences of students and educators, will depend on how well we mobilize knowledge gained from our research, including inviting practitioners into our community and working with them to bridge this gap.

CER is field that is profoundly influenced not only by new developments in chemistry, but also the ever-changing perspectives and values of students and those who support them in classrooms, institutions, and society at large. I believe this idea—that CER, like all types of educational research, is inherently dynamic, complex, and vast—is captured perfectly in a quote from Prof. Joel Westheimer, whose course "Can Education Change the World?" and related talks/papers have greatly influenced my thinking on the role of education and educational research in society (Westheimer, 2015; Westheimer & Kahne, 2004).

> "Education is not rocket science. And by that, I mean it is infinitely more complicated. Orbital speeds and gravitational pull may vary, but [students] vary more. Compared to teaching, rocket science is child's play."

CER is uniquely positioned in that the discipline of chemistry is considered the "central science" that bridges the physical sciences with the applied and life sciences. Chemistry education serves as a fantastic context for which students can learn about chemistry, and also how chemicals can have benefits and hazards across various systems, including environmental, economic, and social systems. Chemistry education can be a venue for equipping students not only with chemistry knowledge, as well as skills and abilities to engage with the world as chemists and scientists—evident in the growth of transformative approaches like Systems Thinking in Chemistry Education (STICE).

All of my work in CER has been guided by my deep desire to help people learn and make learning in chemistry and science more effective and inclusive. I believe I have made intellectual and research contributions to CER in support of these goals, but there is still so much to be done. In the end, I hope that what I have contributed can serve as a starting point for others to contribute, learn, and lead.

12.3 Final reflection

I think many of us hope that the research in our theses will demonstrate our development as researchers and our contributions to important research problems. I believe that my thesis and the work described in it accomplishes this goal.

But I also hope it demonstrates something else.

I grew up in Surrey, British Columbia. If you ask someone from BC about Surrey, it’s likely they’ll make a passive comment about gang and drug violence that has plagued the area for decades. To give readers a sense: Surrey definitely isn’t the most dangerous place in the world, but if you grew up there, it’s likely you know people or friends who ended up in a gang.

My parents raised me in this environment. As immigrants from China, they sacrificed everything and everyone they loved to give my sisters and I a better life than what was possible back home. Science wasn't something I spent time on because I enjoyed it, but because I didn't really have a choice. My parents had already sacrificed so much, and they couldn't have me squandering the opportunity and being caught up in the distractions that caught many of my friends and peers. Perhaps it's because of this upbringing—one that framed science as a *way out* rather than a *way of knowing*—that I've never really felt like a scientist or researcher.

Many of the people I interact with in the scientific community seem to have grown up in environments so foreign to my own. They seem so sure of themselves and their place in science —I see them as *geniuses* who know that this is the life and community for them. I never feel like myself around these people. They appear flow easily between work and home, while I feel like a perpetual outsider.

Yet here I am, writing the final passages of my doctoral thesis. And so I'll dedicate these last few lines for what made this all possible: my mentors.

My mentors have given me space and time to grow. They have offered me empathy and kindness when I have expected punishment. They have given me all the tools I will need to continue learning, publishing, presenting, and excelling as a researcher. They have helped me recognize that my experiences aren't lesser than those "geniuses"—they're just different and too often underappreciated. That I can still do science, even if I don't always feel like a "scientist".

To my fellow first-generation students, immigrants, and "outsiders": Search for mentors who will give you the respect and kindness you need to thrive (not just survive). Search for mentors who will push and challenge you in ways to raise you up, rather than tear you down. Search for mentors who understand that *everyone* has genius inside of them, no matter who they are or where they come from.

Search for mentors like mine: those who believed in me, even when I didn't.

12.4 References for Chapter 12

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *Journal of Chemical Education*, *100*(4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00323

Becker, N., Noyes, K., & Cooper, M. M. (2016). Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *Journal of Chemical Education*, *93*(10), 1713–1724. https://doi.org/10.1021/acs.jchemed.6b00298

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating Students' Reasoning about Acid-Base Reactions. *Journal of Chemical Education*, *93*(10), 1703–1712. https://doi.org/10.1021/acs.jchemed.6b00417

Crossman, E., Lu, Y., & Hou, F. (2022). International students as a source of labour supply: Engagement in the labour market after graduation. *Economic and Social Reports*, *2*(3), 1–5.

Deng, J. M., Carle, M. S., & Flynn, A. B. (2022). Students' Reasoning in Chemistry Arguments and Designing Resources Using Constructive Alignment. In G. V. Shultz & N. Graulich (Eds.), *Student Reasoning in Organic Chemistry*. Royal Society of Chemistry. https://books.rsc.org/books/edited-volume/2035/chapter-abstract/4654521/Students-Reasoning-in-Chemistry-Arguments-and?redirectedFrom=fulltext#:~:text=https%3A//doi.org/10.1039/9781839167782-00074

Deng, J. M., & Flynn, A. B. (2021). Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items. *Chemistry Education Research and Practice*, *22*, 749–771. https://doi.org/10.1039/D0RP00320D

Deng, J. M., & Flynn, A. B. (2023). "I Am Working 24/7, but I Can't Translate That to You":

The Barriers, Strategies, and Needed Supports Reported by Chemistry Trainees from English-as-an-Additional Language Backgrounds. *Journal of Chemical Education*, *100*(4), 1523–1536. https://doi.org/10.1021/acs.jchemed.2c01063

Deng, J. M., Rahmani, M., & Flynn, A. B. (2022). The role of language in students' justifications of scientific phenomena. *International Journal of Science Education*, *44*(13), 2131–2151. https://doi.org/10.1080/09500693.2022.2114299

Dood, A. J., & Watts, F. M. (2022). *Mechanistic Reasoning in Organic Chemistry: A Scoping Review of How Students Describe and Explain Mechanisms in the Chemistry Education Research Literature*. https://doi.org/10.1021/acs.jchemed.2c00313

Emenike, M. E., & Emenike, B. U. (2023). Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. *Journal of Chemical Education*, *100*(4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00063

Herrington, D. G., Sweeder, R. D., Daubenmire, P. L., Bauer, C. F., Bretz, S. L., Bunce, D. M., Carmel, J. H., Dekorver, B. K., Kelly, R. M., Lewis, S. E., Oliver-hoyo, M., Ryan, S. A. C., Stains, M., Towns, M. H., & Yezierski, E. J. (2019). *Supporting the Growth and Impact of the Chemistry-Education- Research Community*. https://doi.org/10.1021/acs.jchemed.8b00823

Higher Education Strategy Associates. (2020). *The state of post-secondary education in Canada*.

Institute of International Education. (2020). *Project Atlas: A Quick Look at Global Mobility Trends*. https://iie.widen.net/s/g2bqxwkwqv/project-atlas-infographics-2020

Jaquette, O., Curs, B. R., & Posselt, J. R. (2016). Tuition Rich, Mission Poor: Nonresident Enrollment Growth and the Socioeconomic and Racial Composition of Public Research Universities. *The Journal of Higher Education*, *87*(5), 635–673. https://doi.org/10.1080/00221546.2016.11777417

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-Language

backgrounds. *Educational Researcher*, *27*(4), 12–21. https://doi.org/10.3102/0013189X027004012

Lieber, L. S., Ibraj, K., Caspari, I., & Graulich, N. (2022). Closing the Gap of Organic Chemistry Students' Performance with an Adaptive Scaffold for Argumentation Patterns. *Chemistry Education Research and Practice*, *2*. https://doi.org/10.1039/d2rp00016d

Martin, P. P., & Graulich, N. (2024). Beyond Language Barriers: Allowing Multiple Languages in Postsecondary Chemistry Classes Through Multilingual Machine Learning. *Journal of Science Education and Technology*, *0123456789*. https://doi.org/10.1007/s10956-023-10087-4

Martin, P. P., Kranz, D., Wulff, P., & Graulich, N. (2023). Exploring new depths: Applying machine learning for the analysis of student argumentation in chemistry. *Journal of Research in Science Teaching*, *August*. https://doi.org/10.1002/tea.21903

McNeill, K. L., & Berland, L. K. (2010). A Learning Progression for Scientific Argumentation: Understanding Student Work and Designing Supportive. *Science Education*, 765–793. https://doi.org/10.1002/sce.20402

McNeill, K. L., & Krajcik, J. (2008a). Inquiry and Scientific Explanations: Helping Students Use Evidence and Reasoning. In *Science as Inquiry in the Secondary Setting* (pp. 121–134).

McNeill, K. L., & Krajcik, J. (2008b). Scientific Explanations: Characterizing and Evaluating the Effects of Teachers' Instructional Practices on Student Learning. *Journal of Research in Science Teaching*, *45*(1), 53–78. https://doi.org/10.1002/tea

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

Moon, A., Moeller, R., Gere, A. R., & Shultz, G. V. (2019). Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students'

writing on ocean acidification. *Chemistry Education Research and Practice*, *20*(3), 484–494. https://doi.org/10.1039/c9rp00005d

Moon, A., Stanford, C., Cole, R., & Towns, M. (2016). The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry. *Chemistry Education Research and Practice*, *17*(2), 353–364. https://doi.org/10.1039/c5rp00207a

National Academies of Science Engineering and Medicine. (2018). *Graduate STEM Education for the 21st Century* (A. Leshner & L. Scherer (Eds.)). National Academies Press. https://doi.org/10.17226/25038

Osborne, J. F., & Patterson, A. (2011). Scientific Argument and Explanation: A Necessary Distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438

Posselt, J. R., & Grodsky, E. (2017). Graduate Education and Social Stratification. *Annual Review of Sociology*, *43*, 353–378. https://doi.org/10.1146/annurev-soc-081715-074324.Graduate

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed From Philosophy of Science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

Ryu, M., Bano, R., & Wu, Q. (2021). Where Does CER Stand on Diversity, Equity, and Inclusion? Insights from a Literature Review. *Journal of Chemical Education*, *98*(12), 3621–3632. https://doi.org/10.1021/acs.jchemed.1c00613

Social Sciences and Humanities Research Council. (2018). *Truth Under Fire in a Post-Fact World*. https://horizons.gc.ca/en/2018/10/19/the-next-generation-of-emerging-global-challenges/#truth-under-fire

Szteinberg, G., Balicki, S., Banks, G., Clinchot, M., Cullipher, S., Huie, R., Lambertz, J., Lewis, R., Ngai, C., Weinrich, M., Talanquer, V., & Sevian, H. (2014). Collaborative professional development in chemistry education research: Bridging the gap between research and practice. *Journal of Chemical Education*, *91*(9), 1401–1408.

https://doi.org/10.1021/ed5003042

United Nations. (2015). *Transforming our World: the 2030 Agenda for Sustainable Development*. https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Watts, F. M., Schmidt-McCormack, J. A., Wilhelm, C. A., Karlin, A., Sattar, A., Thompson, B. C., Gere, R., & Shultz, G. V. (2020). What students write about when students write about mechanisms: analysis of features present in students' written descriptions of an organic reaction mechanism. *Chemistry Education Research and Practice*, *21*, 1148–1172. https://doi.org/10.1039/c9rp00185a

Weinrich, M. L., & Talanquer, V. (2016). Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product. *Chemistry Education Research and Practice*, *17*(2), 394–406. https://doi.org/10.1039/c5rp00208g

Westheimer, J. (2015). No Child Left Thinking. In *What Kind of Citizen? Educating Our Children for the Common Good* (pp. 11–18). Teachers College Press. http://books.google.com/books?hl=en&lr=&id=ZfkglURJOi4C&oi=fnd&pg=PA259&dq=No+Child+Left+Thinking&ots=79jDMjeqR_&sig=4t3VSwVOsz6r2QYEwHMPbp61gOE

Westheimer, J., & Kahne, J. (2004). What kind of citizen? The politics of educating for democracy. *American Educational Research Journal*, *41*(2), 237–269. https://doi.org/10.3102/00028312041002237

Wilson-Kennedy, Z. S., Payton-Stewart, F., & Winfield, L. L. (2020). Toward intentional diversity, equity, and respect in chemistry research and practice. *Journal of Chemical Education*, *97*(8), 2041–2044. https://doi.org/10.1021/acs.jchemed.0c00963

Winfield, L. L., Wilson-Kennedy, Z. S., Payton-Stewart, F., Nielson, J., Kimble-Hill, A. C., & Arriaga, E. A. (2020). Journal of Chemical Education Call for Papers: Special Issue on Diversity, Equity, Inclusion, and Respect in Chemistry Education Research and Practice. *Journal of Chemical Education*, *97*(11), 3915–3918. https://doi.org/10.1021/acs.jchemed.0c01300

Chapter 13: Appendices

13.1 Appendix 1: Chapter 2 Supporting Information

SUPPORTING INFORMATION

Getting past the rules and to the WHY: Causal mechanistic arguments when deciding the plausibility of organic reaction mechanisms

Nicholas E. Bodé, Jacky M. Deng, and Alison B. Flynn*

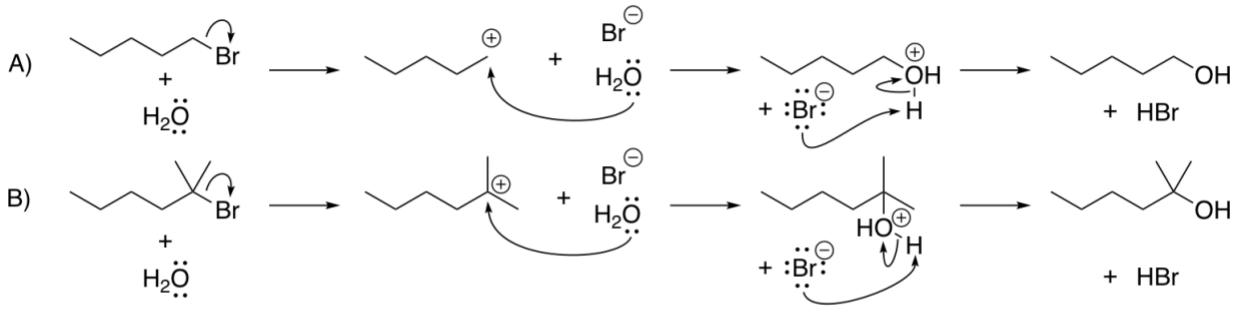
Department of Chemistry and Biomolecular Sciences

Question prompt 2
Coding procedures and Codebook 2
Phase 1: Concepts used, correctly or incorrectly, linking, and comparison 2
Phase 2: What mode of reasoning was found? 4
Coding examples 5
Student 2 (Claim: B) 5
Student 5 (Claim: B) 6
Student 6 (Claim: B) 7
Student 8 (claim: A) 7
Student 10 (claim: B) 8
Student 13 (claim: B) 9
Student 14 (claim: B) 9
Student 22 (Claim: B) 9
Relationship between modes of reasoning and completeness of comparison 10

QUESTION PROMPT

Students' answers to the question in Figure 1 were analyzed in the study.

a. Draw and label the reaction coordinate diagram for the two reactions below that are shown proceeding by an S_N1 mechanism. **(5 points)**



b. Which reaction is most likely to proceed by the mechanism shown? _____ **(1 point)**

c. Justify your answer in part b, using your reaction coordinate diagram as part of your explanation. **(3 points)**

Figure 1. The question analyzed for this study, taken from the students' OC2 final, administered at the end of the fall 2017 semester.

CODING PROCEDURES AND CODEBOOK

Phase 1: Concepts used, correctly or incorrectly, linking, and comparison

This section describes the coding procedure to identify concepts used, how concepts were linked, and whether comparisons were made. Each response was analyzed for five aspects, following the steps below.

1. Is the concept (list in Table 1) discussed (y/n) correctly (c) or with errors (e)?
    - If the discussion of the concept contains no errors, it must be considered correct
    - Any comparisons between reactions A and B should be considered
        - *E.g.*, if the response says that the carbocation in A is more stable than in B, their discussion of this concept would be coded with errors
    - Whether the discussion of that concept is *complete* comes through in other areas of analysis
        - *E.g.*, there will likely be cases where a concept is discussed without errors, but is not linked to any other concepts
    - There may also be cases where the discussion of a concept itself is correct while the link is incorrect
        - For example, consider Student 51's response, which is correct in that the substrate in reaction B is more sterically hindered than in reaction A, but the

link to activation energy is incorrect: "The steric hinderance of the methyl substituents around the bromine leaving group raises the energy of the molecule and also raises the activation energy of the rate determining step, making reaction B less favourable than A. The higher activation energy of the carbocation intermediate in B can be seen in the above diagram."

2. Identify links between concepts
3. For each link between concepts:
    - Is the basis for the link correct (c) or with errors (e)?
        - *E.g.*, answer may relate energy to stability, which makes sense, but the relationship described is incorrect (for instance, saying "1° carbocations are more stable than 3°")
    - Don't code concepts as being linked if you had to go through another concept to get there
        - *E.g.* If an argument has a carbocation (carbo+) linked to stability (stab) which is then linked to activation energy (E_act), do not code carbocation as being linked to activation energy unless there is an explicit link between the two elsewhere.
4. Is a comparison made between the two reactions with respect to that concept (compared reaction A to B)?
    - Comparison can be implicit (*e.g.*, if coding for stability, if they say "the primary carbocation in A is less stable", regard it as them implicitly making a comparison to the tertiary carbocation in B) or explicit.
5. Repeat steps 1–4 for each subsequent concept.

The analysis was organized in an excel spreadsheet as summarized in Figure 2.

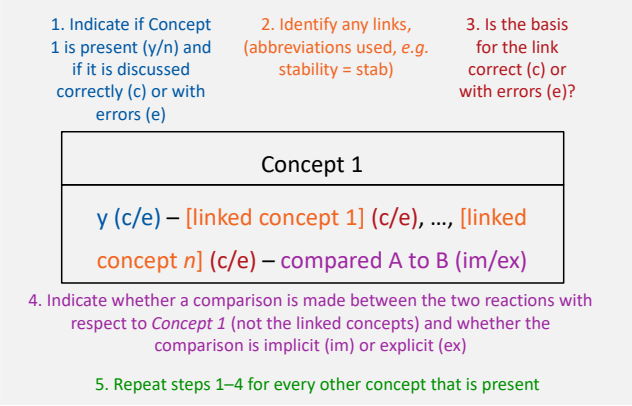


Figure 2. Summary of steps to code for the concepts present, links made, and comparisons made.

Table 1. Codes used for the study with notes to guide their application.

#	Code	Definition	Notes
1	alphaC	Number of α-carbon substituents	Specific reference to the degree of substitution of (or number of methyl groups on) the alpha carbons in the starting materials

			• May include explicit discussion (compare/contrast) regarding presence of Me groups (do not need to explicitly say 1° or 3°, or explicitly say "alpha carbon" if it is clear they are referring to the starting materials and not the carbocation intermediate
2	carbo+	Number of carbocation substituents	Specific reference to the degree of substitution of (or number of methyl groups on the carbocation intermediates • Important to distinguish between when students are discussing the alpha carbon vs. the carbocation intermediate, as it is the latter that holds more relevance to the claim • Student does not explicitly have to say "carbocation" (e.g., "carbon with positive charge" ok) • Do not code unless the answer is SPECIFICALLY talking about the degree of substitution of the carbocation – just mentioning carbocations in general is not a reason for using this code
3	RDS	Rate determining step	Must be mentioned explicitly (i.e., use the words "rate determining step"; "slow step" also accepted)
4	Stab	Stability	Appears to be primary basis for the causal mechanism students provided in their arguments to support B being the correct reaction
5	E_act	Activation energy	The answer must explicitly refer to activation energy • There may be instances where it *seems* like they are referring to E_a (e.g., "energy barrier"), but it must be explicit
6	Rx_like	Reaction likelihood/favourability	Distinct from the claim; would involve general statements about the likelihood/favourability of *any* reaction occurring based on variables such as activation energy, stability of products or intermediates • If a comparison between A and B is made with respect to likelihood, then the statement should be regarded as being a restatement of the primary claim (*i.e.,* would not code rxlike) • Has to be a completely isolated statement from anything related to the primary claim (*e.g.*, student 39: "*the reaction with the lowest E_a is most plausible to happen*")
7	TS	Transition state	Must be mentioned explicitly (*i.e.*, use the words "transition state")
8	E	Energy	• Coded when energy is discussed outside of the context of activation energy ○ *e.g.*, comparing the energy level of the carbocation intermediates in pathways A/B • Could also be coded whenever energy is discussed as part of the answer, but the answer does not make explicit reference to activation energy
9	S_N#	S_N1/S_N2	Distinct from the claim; would involve • Some mention of a SN2 reaction occurring instead of a S_N1 reaction (commonly stated that reaction A would more likely proceed via S_N2 mechanism) • A relational link between 1° alpha carbon/S_N2 and/or 3° alpha carbon/SN1 (*i.e.,* rule-based reasoning) ○ They do not explicitly have to mention both S_N1 and S_N2, if they are saying something like "S_N1 reactions are more likely to occur with 3° alpha carbons [than with 1° alpha carbons]"
10	Sterics	Steric hinderance	Could also include discussion of "bulky" groups or reference to some spatial blocking mechanism • Exception: only makes specific reference to the # of methyl groups on the alpha carbons in the starting materials (code 1) without further discussion Note that they might say B is more sterically hindered (which is correct), but the error (sterics are irrelevant to this question) will come out of the data elsewhere (links to other concepts, description of the overall argument) • If they say B is more sterically hindered than A, code this as correct
11	HypCon	Hyperconjugation	Could include similar statements that provide a justification for carbocation stability that is based on the electronic influence of neighbouring alkyl groups • *i.e.,* any canonical description of hyperconjugation that does not include the word "hyperconjugation" is acceptable • element of electronics is key
12	Other	Other	Provide supplementary comments regarding what the "other" is referring to • If this "other" concept/theme is found to be recurring, add it to the codebook

Phase 2: What mode of reasoning was found?

Each argument was analyzed to determine if it included **a descriptive, relational, or causal argument** (Table 2). For example, does the answer explicitly relate energy and stability (is there a "because…" statement to support the link)?

Table 2. Comparison between descriptive, relational, and causal explanations.

Mode of reasoning	Description
Descriptive	Claims are supported solely by describing the properties of the reaction materials (starting materials, intermediates, products). Explicit (surface) features of the problem are described without further explanation.
Relational	Explicit and implicit properties of the reaction materials are highlighted and discussed; comparisons may be made between the materials in each reaction presented Relationships between properties of the reaction materials and their activities are discussed in a correlative fashion (i.e., a cause-effect relationship is not established)
Linear causal	All of the features of a relational argument are present, and are accompanied by cause-and-effect relationships between the relevant properties of the reaction materials and their activities Properties do not necessarily have to be relevant to a canonical explanation, but they must be relevant to the argument the student has provided
Multi-component causal	Phenomena are seen as the result of the static or dynamic interplay of more than one factor and the direct inter- actions of several components. Causal stories are built. Explicit and implicit features of a system are noticed. Isolated: effects of several variables are considered and weighed separately. Integrated: arguments as interconnected stories of how different variables affect the entities involved.

Concepts in a causal argument may not be perfectly connected to one another, but must include a causal influence for why one reaction is more or less likely to occur, or for why one reaction is higher/lower in activation energy.

The nature of the link between each concept was not described because multiple concepts (at least three) typically need to be woven together to provide a sophisticated argument (the example of a causal argument above contains three inter-related concepts). As a result:

- The mode of reasoning in the argument was only applied to the argument as a whole
- Some portions of the argument may be at higher/lower level than other portions; therefore, the coding describes only the highest level of argument that is provided (regarding the argument in its entirety)

CODING EXAMPLES

Student 2 (Claim: B)

> *"B is more likely to proceed by the mechanism shown since S_N1 reactions don't typically happen on 1° carbons like in A"*

Alpha carbon substitution (1)	S_N1/S_N2 (9)
y (c) - S_N# (c)	y (c) - alphaC (c)

Figure 3. Representation of coding for Student 2's answer in Excel spreadsheet. Overall, this is a ***descriptive*** explanation; the statement is correct but invokes a memorized rule with no **causal**

mechanism (*i.e.*, a reason is not provided for **why** S_N1 mechanisms do not typically occur on 1° carbons). The concepts discussed include α carbon substitution and S_N1 reactions (Figure 3, top 2 boxes); these concepts are correctly linked (Figure 3, bottom two boxes). No explicit comparison is made between reaction mechanisms A and B; Mechanism B's feasibility is not discussed explicitly.

Student 5 (Claim: B)

> *"An initial starting material of a 3° carbon next to the LG [leaving group] is favoured since it forms that 3° stable enough carbocation, a 1° carbon will not make a stable enough carbocation to form"*
>
> *… a 1° carbon will not make a stable enough carbocation to form, so pathway A would have been a concerted SN2 mechanism with no carbocation formation step*
>
> *… the creation of the "A" carbocation is unlikely since it requires a lot more energy (higher activation energy) to form and thus is very unstable*
>
> *… reactions favour paths of less energy input (where it is easier to form products)"*

Alpha carbon substitution (1)	Carbocation (2)	Stability (4)	Activation Energy (5)	Reaction likelihood/ favourability (6)	Energy (8)	S_N1/S_N2 (9)
y (c) - Carbo+ (c), S_N# (c), Stab (c), Compared A to B	y (c) - AlphaC (c), Stab (c), S_N# (c), Compared A to B	y (c) - alphaC (c), E_act (c), Compared A to B	y (c) - Stab (c), Compared A to B	y (c) – E (c)	y (c) – Rx_like (c)	y (c) - AlphaC (c), Carbo+ (c), Compared A to B

Figure 4. Representation of coding for Student 5's answer in Excel spreadsheet.

Correct links between 1 and 2, 2 and 4; comparisons between A and B for 1, 2, and 4.

> *"…a 1° carbon will not make a stable enough carbocation to form, so pathway A would have been a concerted S_N2 mechanism with no carbocation formation step"*

Correct link between 2 and 9, comparison between A and B for 9

> *"the creation of the "A" carbocation is unlikely since it requires a lot more energy (higher activation energy) to form and thus is very unstable"*

Correct link between 4 and 5, comparison between A and B for 5 (because they say formation of carbocation in A requires <u>more</u> energy)

> *"reactions favour paths of less energy input (where it is easier to form products)"*

Correct link between 6-8

Overall, this is a **linear *causal* explanation**; stability is invoked to explain the cause-and-effect relationship between carbocation substitution and activation energy.

Student 6 (Claim: B)

> *"…the tertiary carbocation formed as intermediate in B is more stable than the primary carbocation formed in A. As a result, the activation energy of B is lower than that of A.*
>
> A reaction that consumes a lot of energy to proceed is least likely to occur than that consume less energy since the energy barrier is lower."

Correct link between 2 and 4, comparison between A and B for 2 and 4 (Figure 5): **"**…the tertiary carbocation formed as intermediate in B is more stable than the primary carbocation formed in A."

Correct link between 4 and 5, comparison between A and B for 5: **"**As a result, the activation energy of B is lower than that of A"

Correct link between 6 and 8: "A reaction that consumes a lot of energy to proceed is least likely to occur than that consume less energy since the energy barrier is lower."

Overall, this is a ***linear causal*** explanation; stability is invoked to explain the cause-and-effect relationship between carbocation substitution and activation energy.

Carbocation substitution (2)	Stability (4)	Activation Energy (5)
y (c) - stab (c), E_act (c), Compared A to B	y (c) - carbo+ (c), E_act (c), Compared A to B	y (c) - Carbo+ (c), Stab (c), Compared A to B

Figure 5. Representation of coding for Student 6's answer in Excel spreadsheet

Student 8 (claim: A)

> "Reaction B is more sterically hindered than reaction A. This means that it would take more energy to reach the activation energy in reaction B than in reaction A."

Concept 5 is discussed incorrectly and concept 10 is discussed correctly; B is more sterically hindered; however steric hindrance is not relevant to the activation energy for this mechanism (Figure 6).

Overall, this is a ***linear causal*** argument (with errors). Steric hinderance provides a causal link (although not correctly) to the relative activation energy barriers relevant to their claim.

Activation Energy (5)	Reaction likelihood/favourability (6)	Energy (8)	Steric hinderance (10)
ye - Sterics (e), Compared A to B	yc, E (c)	yc, Rx_like (c)	yc – E_act (e), Compared A to B

Figure 6. Representation of coding for Student 8's answer in Excel spreadsheet

Student 10 (claim: B)

> *"Reaction B experiences greater steric hinderance, and referring to the energy diagram that means it requires a much higher activation energy than reaction A which has little steric hinderance, and by looking at the diagram we can see this causes lower activation energy. For this reason, reaction B is more likely to proceed using the 2 step S_N1 mechanism as shown in reaction B, whereas B [Meant A?]is more likely to proceed using S_N2 which requires a lower E_a"*

Incorrect link is made between concepts 5 and 10; comparison made between A and B for 5 and 10. Concept 5 is discussed incorrectly but concept 10 discussed correctly (true that B is more sterically hindered, it just is not relevant): "Reaction B experiences greater steric hinderance, and referring to the energy diagram that means it requires a much higher activation energy than reaction A which has little steric hinderance, and by looking at the diagram we can see this causes lower activation energy."

Incorrect link is made between concepts 5 and 9; concept 9 is discussed incorrectly (likely meant to say A instead of B for one of those statements): "For this reason, reaction B is more likely to proceed using the 2 step S_N1 mechanism as shown in reaction B, whereas B [meant A?] is more likely to proceed using S_N2 which requires a lower E_a."

Overall, this is **a linear causal** argument (with errors).

Activation Energy (5)	S_N1/S_N2 (9)	Steric hinderance (10)
ye - Sterics (e), S_N# (e), Compared A to B	ye – E_act (e)	ye – E_act (e), Compared A to B

Figure 7. Representation of coding for Student 10's answer in Excel spreadsheet

Student 13 (claim: B)

> *"B will proceed because even though it requires greater activation energy it forms the most stable product because it forms the most substituted product."*

Incorrect link between concepts 1 and 4 (Figure 8). Concept 1 is discussed correctly; concept 4 is discussed incorrectly. Concept 5 is discussed incorrectly. Overall, this is a **relational** argument (with errors).

Alpha carbon substitution (1)	Stability (4)	Activation Energy (5)
ye - Stab (e), Compared A to B	ye - AlphaC (e), Compared A to B	ye - Compared A to B

Figure 8. Representation of coding for Student 13's answer in Excel spreadsheet.

Student 14 (claim: B)

> "The reaction B is more likely to occur than reaction A because the alpha carbon is a tertiary carbon. This means that it is more reactive and requires less energy for the reaction to occur. This means that the activation energy is lower and easier to reach. Also, an S_N1 reaction prefers a 3° carbon to occur.
>
> *In contrast, reaction A contains a primary alpha carbon. This alpha carbon is not as reactive as an S_N1 reaction would need it to be and therefore it will require much more energy to be carried out. The activation energy will be higher and harder to react. The primary alpha is more suitable for an E2 or S_N2 reaction."*

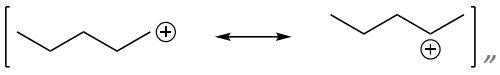
Links are made between concepts 1 and 5, 1 and 9, and 5 and 9; all discussed correctly (Figure 9). The argument is **relational**; the reason for each of the links between concepts are explained (*e.g.*, why an "S_N1 reaction prefers a 3° carbon to occur").

Alpha carbon substitution (1)	Activation Energy (5)	SN1/SN2 (9)
yc - S_N# (c), E_act (c), Compared A to B	yc - AlphaC (c), Compared A to B	yc - alphaC (c), Compared A to B

Figure 9. Representation of coding for Student 14's answer in Excel spreadsheet.

Student 22 (Claim: B)

> *"S_N1 reactions heavily favour third degree carbocations over primary carbocations. In the case of A, the activation energy to remove bromine and form a 1° carbocation is very high. In B, the activation energy is much lower to produce a 3° carbocation. Reaction A does not occur at all. The carbocation would shift to a 2° carbocation though proton transfer [drawing]. Thus B is much more likely to proceed by the shown mechanism.*

Drawing:  „

Carbocation substitution	Rate determining step	Activation Energy	SN1/SN2	Other (say what it was in comments)
yc – S_N# (g), E_act (c), Compared A to B	y (c)	y (c) - carbo+ (c), Compared A to B	y (c) - carbo+ (c)	y (c)

Figure 10. Representation of coding for Student 22's answer in Excel spreadsheet

This is a ***relational*** explanation. The answer illustrates how a multicomponent answer where none of the discussion is really *incorrect* can still fail to properly provide a causal mechanism for **why** B is more likely than A to proceed by the mechanism shown:

- First statement cites a rule
- Second statement invokes activation energy differences
- Correct that E_a barrier to a 1° carbocation is higher, but they don't explain why

RELATIONSHIP BETWEEN MODES OF REASONING AND COMPLETENESS OF COMPARISON

The mode of reasoning was not a factor in the completeness of the comparison; a similar distribution in comparisons was found across modes of reasoning for arguments with correct and incorrect claims.

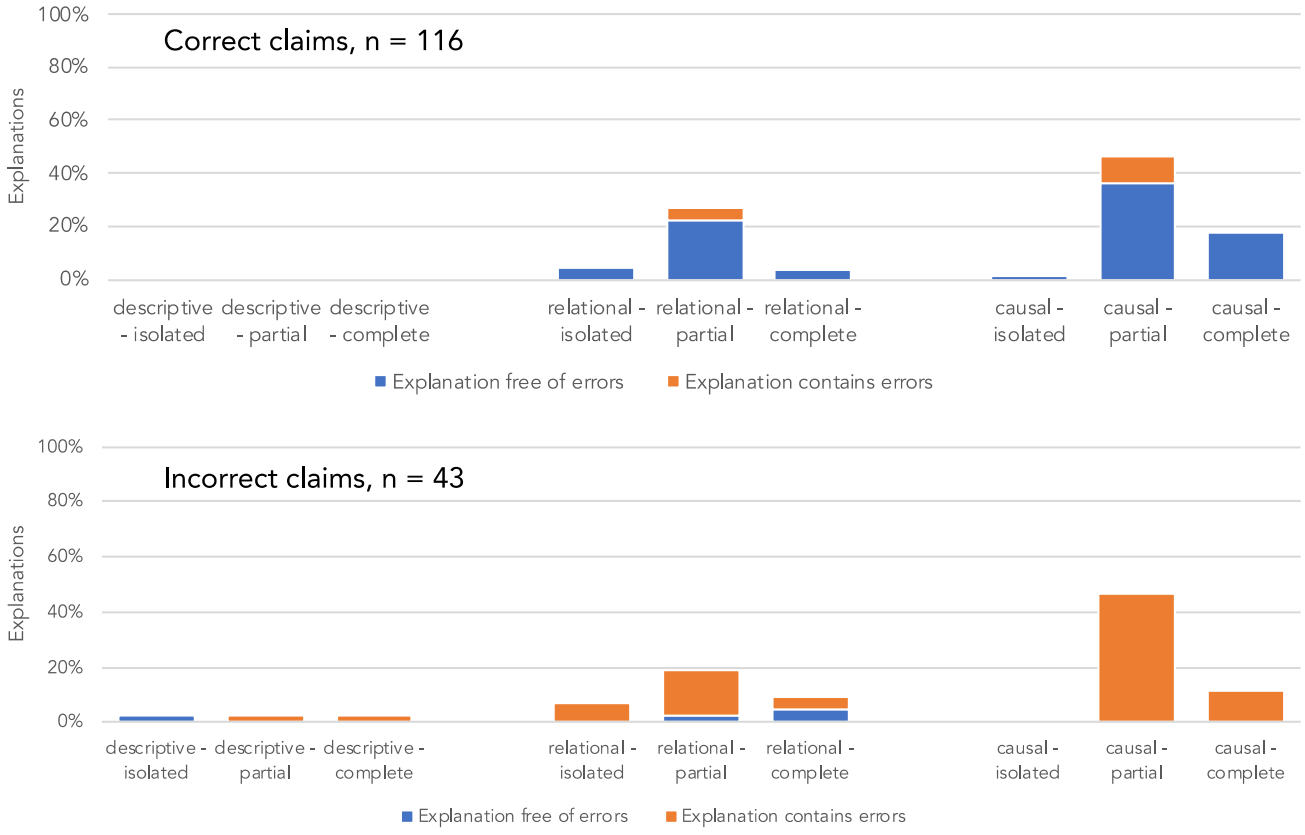


Figure 11. Degree of comparison of concepts (isolated, partial, complete), organized by mode of reasoning. N = 159.

13.2 Appendix 2: Chapter 5 Supporting Information

Supplementary Materials: The role of language in students' justifications of scientific phenomena

Jacky M. Deng[1], Malek Rahmani[1], Alison B. Flynn[1]*
Correspondence to: alison.flynn@uottawa.ca

Table of Contents

Methods and additional materials ***2***

Ocean acidification argumentation prompt **8**

Validity evidence for Language Experience and Proficiency Questionnaire (LEAP-Q) **8**

Study design **9**

Qualitative analysis of students' arguments **9**

Codebook for characterizing chemistry students' arguments about ocean acidification ***11***

Example of multi-component causal argument: **14**

Reasoning diagrams for each mode of reasoning **15**

Inter-rater reliability of qualitative coding **16**

Generating language profiles from LEAP-Q data ***17***

Statistical analyses ***17***

Fischer's exact test **17**

Cluster analysis **17**

Logistic regression **17**

Follow-up interviews ***17***

Follow-up Interviews Protocol **19**

Supplemental Text ***22***

Argumentation in education **22**

English learners in STEM **22**

Fig. S1. Distributions of English proficiency and history for Eng+ participants (n = 28, top) and Eng1st participants (n = 138, bottom) ***24***

Fig. S2. Reasoning demonstrated in responses per academic year ***25***

Table S1. Descriptions of modes of reasoning and exemplars from dataset ***26***

Table S2. Collinearity statistics for logistic regression with English proficiency and history (predictors) and reasoning. ***27***

Table S3. Logistic regression output. Dependent variable: mode of reasoning. Independent predictors: English history, English proficiency, academic years 1-4. ..... ***28***

Table S4. Logistic regression output. Dependent variable: mode of reasoning. Independent predictors: Time in English language school, time in English language home, and time in English language country.....29

Table S5. Summary of follow-up interview participants (n = 16). Participants had the option to select their own pseudonym and participants who did not have a preference were assigned a pseudonym by the researchers. Pseudonyms assigned by researchers were selected to ensure anonymity while also considering the identities expressed by interviewees during interviews (e.g., cultural/language background).....30

Table S6. Participants' perspectives on supporting students from non-dominant language backgrounds, including suggested resources and/or ways to support. 31

Table S7. Eng1st and Eng+ participants' perceptions on how context influences how they respond. 32

Table S8. Participants' perceptions on how their language backgrounds contributed to how they constructed written arguments in chemistry. 33

Table S9. Output from K-means cluster analysis using four-cluster solution.....34

Table S10: Summary of interview themes and quotes (see attached file).....35

References.....36

Methods and additional materials

Full instrument – hosted on SurveyMonkey

Invitation to participate: I am invited to participate in the above-mentioned research study conducted by Jacky Deng (PhD student), co-investigator (CI), in partial fulfillment of the requirements for his PhD thesis, and Professor Alison Flynn, the principal investigator (PI).

Purpose of the study: The purpose of this study is to examine science students' abilities to construct arguments from evidence. Specifically, the goal is to explore students' abilities to construct arguments from evidence about climate change issues.

Participation: To participate, I will complete one online survey. The survey will include providing demographic information and responding to a question about a climate change scenario. I may also choose to provide my email to the researchers in case they have questions about my responses.

Risks: Participants' names, including mine, will be kept by the PI and CI. In the event that the PI or CI is teaching or involved in the course, participant names will be kept by a research assistant and will only be released to the PI or CI after the course marks are finalized. I can stop the survey at any time and can consult SASS if needed (sass.uottawa.ca). I have received an assurance from the researcher that every effort will be made to minimize these risks, such as maintaining confidentiality (i.e., using a pseudonym), and keeping the files on a password protected computer and any hard copies in a locked office.

Benefits: I will be contributing to improving the quality of science education for myself and my peers. I might also gain awareness of my own knowledge and values about climate change.

Confidentiality and anonymity: I have received assurance from the researcher that the information I share will remain strictly confidential. The researchers will NOT share the names of any participant with anyone

else at any time. Referring to me using a pseudonym in all communications, including any quotes, will ensure anonymity.

Conservation of data: All digital data will be kept in PI and CI's password protected computer or in a locked cabinet in a locked office or storage room. The PI will have access to the data once course grades have been finalized, while the CI will have access the entire time. The data will be stored for 15 years after completion of the project (i.e., 2035), after which it will be destroyed and deleted.

Compensation: To thank you for your contribution to the research project: you will be given the option to enter your name in a draw to win an Amazon gift card valued at $20 to $100, dependent on the number of participants in the study.

Voluntary participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions, without suffering any negative consequences. Choosing to participate (or not) will have no effect on my academic standing or relationship with the professor conducting the project. If I choose to withdraw, all data gathered up to the time I choose to withdraw will be removed from the project.

By proceeding, I am consenting to participate in the above research study conducted by Jacky Deng and Professor Alison Flynn, by completing an online survey. If I have any questions about the study, I may contact Jacky Deng ([redacted]) or Alison Flynn ([redacted]). If I have any questions regarding the ethical conduct of this study, I may contact the protocol officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5. Tel: (613) 562-5387; Email: ethics@uottawa.ca

1. In which course did you hear about this survey? Please enter the course code (*e.g.*, CHM 2120). **[open-response]**

2. Which other chemistry courses are you currently enrolled in or have you previously completed? **[checkboxes]**

3. What level of education are you currently in? **[multiple choice]**

    - 1st
    - 2nd
    - 3rd
    - 4th or higher
    - Other (please specify)

4. Is English one of your first languages? A first language is defined as a language you have learned since birth and continue to use today. **[multiple choice]**

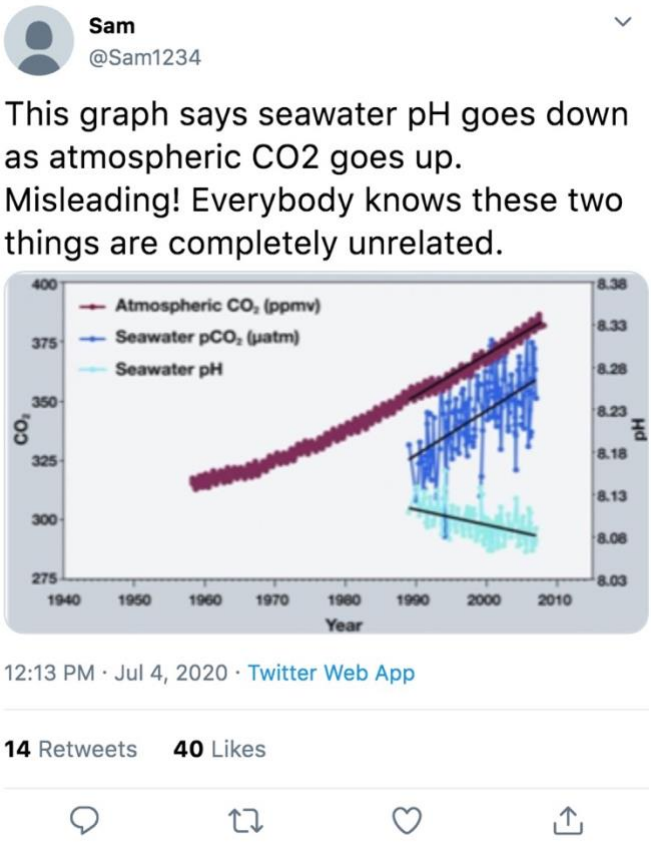
    - Yes
    - No

5. What is the main language you have learned chemistry in? **[open-response]**

6. If you would like to be included in the draw for an Amazon gift card, please list your email and name: **[open-response]**

7. If English is not your first language, the following two questions will be the same, but you will be asked to respond in either your first language or English.

Consider the following social media post:



Also consider this series of chemical equilibria and pK_a values:

$$CO_{2(aq)} + H_2O_{(l)} \rightleftharpoons H_2CO_{3(aq)}$$

$$H_2CO_{3(aq)} + H_2O_{(l)} \rightleftharpoons HCO^-_{3(aq)} + H_3O^+_{(aq)}$$

$$HCO^-_{3(aq)} + H_2O_{(l)} \rightleftharpoons CO^{2-}_{3(aq)} + H_3O^+_{(aq)}$$

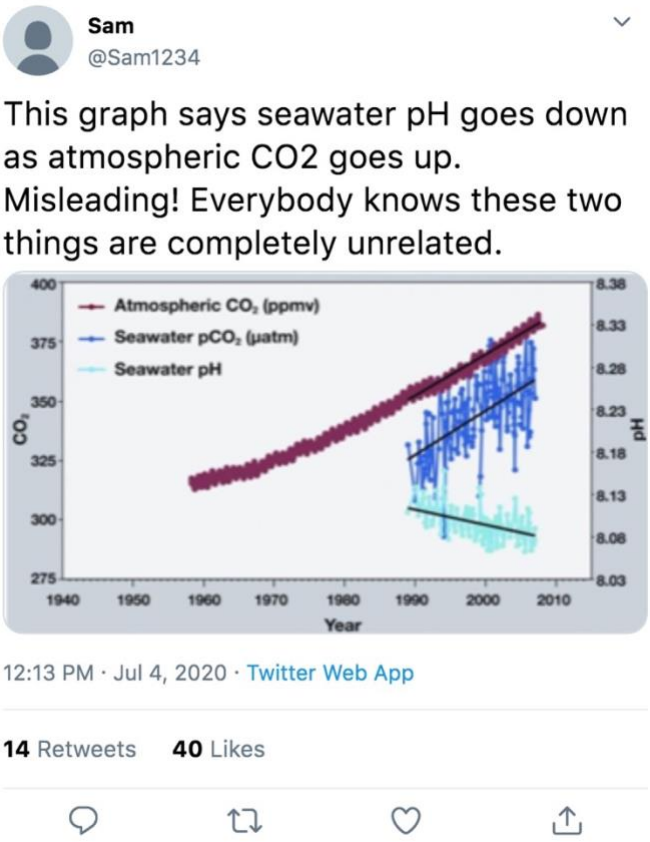
$$pK_a\ H_2O_{(l)} = 14.0$$
$$pK_a\ H_2CO_{3(aq)} = 6.35$$
$$pK_a\ HCO^-_{3(aq)} = 10.3$$

Given this chemical information, do you agree or disagree with Sam? Justify your decision in a written argument of at least 3 sentences. You are welcome to use the chemical equations to help construct your response, along with any of your own knowledge and experience that you think is relevant. Please respond in English. **[open-response]**

8. What is your non-English first language? **[open-response]**

9. Your non-English response to the next question will be translated into English during our analysis. Would you be interested in being contacted by the researchers in case they have questions about your response during the translation process? If so, please enter your email below. **[open-response]**

10. Consider the following social media post:



Also consider this series of chemical equilibria and pK_a values:

$$CO_{2(aq)} + H_2O_{(l)} \rightleftharpoons H_2CO_{3(aq)}$$

$$H_2CO_{3(aq)} + H_2O_{(l)} \rightleftharpoons HCO^-_{3(aq)} + H_3O^+_{(aq)}$$

$$HCO^-_{3(aq)} + H_2O_{(l)} \rightleftharpoons CO^{2-}_{3(aq)} + H_3O^+_{(aq)}$$

$$pK_a\ H_2O_{(l)} = 14.0$$
$$pK_a\ H_2CO_{3(aq)} = 6.35$$
$$pK_a\ HCO^-_{3(aq)} = 10.3$$

Given this chemical information, do you agree or disagree with Sam? Justify your decision in a written argument of at least 3 sentences. You are welcome to use the chemical equations to help construct your response, along with any of your own knowledge and experience that you think is relevant. Please respond in **your first language.** Feel free to use English/French chemistry terms, if needed. **[open-response]**

11. If you would rather handwrite your response, attach a file now if you wish to upload an image or PDF file of your non-English response. **[file upload]**

12. How did you feel about responding in your first language? **[open-response]**

13. Please list all the languages you know **in order of dominance: [open-response]**

14. Please list all the languages you know **in order of acquisition**. **[open-response]**

15. Please list what percentage of time you are *currently* and *on average* exposed to each language. Your percentages should add up to 100%. **[dropdown boxes, 5% intervals]**

16. When choosing to read a text available in all languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. Your percentages should add up to 100%. **[dropdown boxes, 0-100%, 5% intervals]**

17. When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you speak in each language? Please report a percent of total time. **[dropdown boxes, 0-100%, 5% intervals]**

18. Age when you... **[open-responses]**

    - began acquiring English
    - became fluent in English
    - began reading in English:
    - became fluent reading in English:

19. Please list the number of years and months you spent in each language environment. **[dropdown boxes (years, 0-30; month, 1-11)]**

    - A country where English is spoken:
    - A country where English is spoken:
    - A country where English is spoken:

20. On a scale from 0 to 10, please select your *level of proficiency* in speaking, understanding, reading, and writing in English:

	0 - none	1 – almost none	2 – very low	3 - low	4 – slightly less than moderate	5 - moderate	6 – slightly more than moderate	7 - good	8 – very good	9 - excellent	10 - perfect
Speaking											
Understanding											
Reading											
Writing											

21. On a scale from 0 to 10, please select how much the following factors contributed to you learning English:

	0 – not a contributor	1 – minimal contributor	2	3	4	5 – moderate contributor	6	7	8	9	10 – most important contributor
Interacting with friends											
Interacting with family											
Reading											
Apps/websites/self-instruction											
Watching TV											
Listening to podcasts/music											

22. Please rate to what extent you are currently exposed to English in the following contexts:

	0 – never	1 – almost never	2	3	4	5 – half of the time	6	7	8	9	10 – always
Interacting with friends											
Interacting with family											
Reading											
Apps/websites/self-instruction											
Watching TV											
Listening to podcasts/music											

Ocean acidification argumentation prompt

We developed an instrument to collect (1) participants' arguments in a chemistry context and (2) participants' language proficiencies and histories. The first part of the instrument prompted participants to construct an argument in agreement/disagreement of a social media claim denying data related to ocean acidification. We decided on an ocean acidification context for several reasons: (1) a similar prompt had been used in previous work in chemistry education research to effectively elicit participants' argumentation (Alena Moon, Moeller, Gere, & Shultz, 2019), (2) the chemical mechanisms behind ocean acidification are well-established and rely on concepts from introductory courses (acid–base equilibria, Le Chatelier's principle), and (3) ocean acidification is a consequence of global warming, allowing us to explicitly probe participants' perspectives on a complex global issue.

Validity evidence for Language Experience and Proficiency Questionnaire (LEAP-Q)

For the second part of the instrument, we relied on the established Language Experience and Proficiency Questionnaire (LEAP-Q) (Kaushanskaya, Blumenfeld, & Marian, 2019; Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAP-Q asks respondents to self-report their language background and perceived proficiency and history with specific languages (in this study, English). The LEAP-Q has been used extensively as a way to measure language proficiency/history in respondents in cases where full-length language tests are not feasible (Dimitropoulou, Duñabeitia, & Carreiras, 2011; Krizman, Skoe, Marian, & Kraus, 2014; Reichle & Birdsong, 2014), with evidence for validity based on relationships with behavioural measures of language proficiency/history and internal structure (Marian et al., 2007).

Prior to administering the instrument, we conducted response process interviews to evaluate response-process validity evidence elicited by the instrument's (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 2014). Response process was the primary threat to validity in this context given that students from diverse language backgrounds were being asked to interpret and respond to items in English (Willis, 2011; Zumbo & Hubley, 2017). We conducted five rounds of cognitive interviews with a total sample of 25 interviewees. The protocol's initial design was based on probing respondents' interpretations and responses to items we anticipated to be linguistically challenging and/or ambiguous (Sorana Laneanu, 2016; Zumbo & Hubley, 2017). We were concerned that the initial version of the argumentation prompt might be too linguistically complex for some students, so a portion of questioning during interviews was focused on investigating how students interpreted the question (meaning and clarity of wording), as well as how students generated their response. Other discussions focused on text labels for Likert scales in the LEAP-Q, as the challenges of establishing validity with Likert scales are well-reported (Barbera & VandenPlas, 2011; DeCastellarnau, 2018; Jamieson, 2004). In interviews, we provided interviewees with Likert text labels in a randomized order, and then asked them to order the terms from lowest to highest. Instrument modifications were made after the initial round of interviews. Following the initial round, subsequent rounds of interviews were dedicated to (1) investigating whether modifications had alleviated response process issues identified by the researchers in previous rounds, and (2) identifying any additional threats to respondents' response processes. Interviewing continued until no additional issues were

revealed in additional rounds. Modifications that were made to the instrument over the course of the interviews were focused on (1) simplifying language throughout the instrument, which a specific focus on items that interviewees described as confusing and/or ambiguous, (2) updating text labels and their organization on the LEAP-Q's Likert scales to better reflect interviewees' interpretations of scale labels.

Study design

Students were invited from chemistry courses at the University of Ottawa to complete the instrument in Fall 2020 and Winter 2021. A recruitment email was sent to instructors for each course describing the study and with a SurveyMonkey link to the instrument. A total of 166 participants submitted a complete response for the study (*n* = 28 Eng+; *n* = 138 Eng1st).

Qualitative analysis of students' arguments

We analyzed participants' arguments using the modes of reasoning framework (Bodé, Deng, & Flynn, 2019; I. Caspari, Kranz, & Graulich, 2018a; Deng & Flynn, 2021; A. Moon, Stanford, Cole, & Towns, 2016; M. Weinrich & Talanquer, 2016). This framework has been previously used to characterize participants' arguments and identifies the type of reasoning participants engage in when explaining or arguing about chemical phenomena.

Descriptive arguments list or give features and/or the properties of entities without establishing connections. Descriptive arguments in this work were characterized as arguments in which the data and/or equilibria provided were described at a surface-level with no explicit connections established between the two (*e.g.*, stating that seawater pH goes down as atmospheric CO_2 levels increase).

Relational arguments include connections between properties of the entities and their activities, but these relationships are discussed in a correlative fashion (*i.e.*, absent of causality). In other words, connections are stated but the argument does not extend to *why* these links or evidence are appropriate. Relational arguments in this work were characterized as arguments in which the data and equilibria provided were described and related to each other (*e.g.*, connecting the pH/CO_2 relationship to the relevant chemical equilibria).

Causal arguments include all features of a relational argument and additionally contain cause-and-effect relationships between the relevant properties of the entities and their activities. Links are stated and additional reasoning explains why or how these links are relevant and/or appropriate, often by referencing scientific knowledge, principles, additional evidence, *etc*. *Linear causal* arguments establish a single chain of causal relationships between one or more pieces of evidence to justify a claim. Linear causal arguments in this work were characterized as arguments in which the link between the data and equilibria justified a causal relationship with a chemical concept (*e.g.*, using Le Chatelier's principle to demonstrate how increasing atmospheric CO_2 causes decreased seawater pH). *Multi-component causal* arguments establish multiple chains of causal relationships between more than one piece of evidence to support a claim. Multi-component causal arguments are similar to linear causal arguments and also include more than one chain of causal reasoning to support the claim being advanced (*e.g.*, using both Le Chatelier's principle and pK_a data to justify). Following analysis, linear causal and

multi-component causal responses were combined into a single group (causal) because only three responses demonstrated multi-component causal reasoning.

We developed a coding protocol based on the features that participants discussed in their arguments. For example, the code *graph* was developed to capture participants' explicit descriptions of the graph in the prompt. Other codes, such as *equations_general* and *equation1* allowed us to capture statements related to the chemical equilibria provided to participants, while also allowing us to differentiate between participants who discussed the equilibria generally ("the equations show…") or more specifically ("the first equation shows…").

Codebook for characterizing chemistry students' arguments about ocean acidification

For each argument, identify the claim being made and the evidence and concepts being used to justify the claim:

Student response	Claim	CO2_location
I disagree with Sam. As atmospheric CO_2 increases…	D	X

Determine where the argument agrees (A) or disagrees (D) with Sam's position.

Code in spreadsheet as:	Claim	Notes and examples
A	Agrees with Sam	**"I somewhat agree with Sam.** With the equations, I don't think a direct link can be made as I believe a high Ka means the acid won't dissociate as much." - Some responses in which the student says they agree but provide an argument in disagreement (i.e., misunderstanding about Sam's position). These should be coded as **disagree** because the argument as a whole is in disagreement with Sam's claim.
D	Disagrees with Sam	"**I disagree with Sam** because according to the chemical equations, carbon dioxide and water form carbonic acid which has a lower pH than water…"

Determine whether the argument discusses (X) or does not discuss (blank) specific pieces of evidence or concepts.

Label in spreadsheet	Evidence/concept	Notes and examples
graph	Explicit callout to graph	**"Based on the graphical data shown alone**, it already appears that there is a clear correlation between atmospheric CO2 levels and the decrease of seawater pH levels…" - Goal of this code is to capture students' acknowledgement of the graphical data - **Do not code** if argument states information present in the graph *without* explicitly stating that they are using the graph to get this information. For example, "there is a clear correlation between atmospheric CO2 levels and the decrease of seawater pH levels" would not count as an explicit callout to the graph. Cannot assume that the student is using the graph to get this information or some other prior knowledge
CO2_location	Specific reference to CO2	"**As atmospheric CO_2 increases** there is more CO2 to interact and react with the water on the surface of the ocean…" - Goal of this code is to capture students' acknowledgement that either atmospheric aqueous CO_2 levels are relevant to the claim - Code for any references to *atmospheric* or *aqueous* CO_2 levels
CO2_general	General reference to CO2	"**As CO2 increases**, the equilibria are driven to products…" - Goal of this code is to capture students' general references to CO_2 - Code if the response discusses CO_2 levels, but without specification of where and/or what state (aqueous or atmospheric) - Do not code if discussed in the context of equation 1 (in those cases, code as equation1).
pH/acidity	Decrease in seawater pH or increase in seawater acidity	"This shift ultimately increases the number of Hydronium (H+) ions. If H+ increases, **then the pH decreases**, in other words, the **after [sic] becomes more acidic**…" - Goal of this code is to capture students' acknowledgement that a decrease in pH and/or increase in acidity is relevant to the claim - Code for any references to seawater pH decrease or acidity increase. - Often may not be explicit about seawater (like in example). Still code in these cases as it is implied that it is the seawater pH that is changing. - **If pH is discussed incorrectly ("pH goes up") or mistaken with pK_a, codes as Y**
equations_gen	General to equations/equilibria	"**Based on the chemical equations provided** above and the fact that our current climate is following a similar trajectory, I strongly disagree with Sam…" - Goal of this code is to capture students' acknowledgement of the equilibrium data - Terms that warrant a code: mechanisms, equilibria, equations, formula - Code for this **even if the argument elaborates on specific equations later on**. For example: "Based on the equations, the increased

		amount of CO2 is reacting with water to form H2CO3" would be coded as both an explicit callout and explicit discussion of equation 1.
equation1	Explicit discussion of equation 1 forming acid	"**The CO2 reacts with water to form carbonic acid (H2CO3).** This carbonic acid then reacts with the water to form bicarbonate (HCO3), while releasing H+ ions (which form H3O+ ions with water)..." - Goal of this code is to capture students' discussion for the first equilibrium shown
equation23	Explicit discussion of equations 2 and/or 3 forming acid	**"**While CO2 as a substance in isolation may not be considered especially acidic, its reaction upon contact with water, as per the above reaction mechanism, yields carbonic acid. **The following two reaction mechanisms once again yield H3O+ which is an Arrhenius acid…"** - Goal of this code is to capture students' discussion for the second two equilibria shown - Often these two equilibria are lumped together in students' arguments. However, some also only discuss one of the two. For both situations, use this code.
LeChat_state	States Le Chatelier's principle only (does not define or explicitly apply)	**"**Although the equilibrium of the second equation lies on the products side, as the amount of CO2 increases, **the amount of H3O+ will increase by Le Chatelier's Principle..."** - Goal of this code is to capture students' responses that cite Le Chatelier's principle, but may be doing so in a rote fashion
LeChat_apply	Defining and/or applying Le Chatelier's principle	"Carbon dioxide will react with the water to form carbonic acid in an equilibrium. **This equilibrium is disrupted by the increase in carbon dioxide concentration which drives the equilibrium to form more carbonic acid as per Le Chatelier's principle…**" - Goal of this code is to capture students' responses that define or Le Chatelier's principle to justify how increasing CO_2 influences the equilibria - Can be applied in cases where it is clear that the student is applying through using words like "push", "shifts", "drives" - Does not have to state "Le Chatelier's principle"
pKa	Uses pK_a values when determining direction of equilibria	**"Carbonic acid has a pk_a of 6.35, meaning that most of it will be in its deprotonated form at a pH > 6.35**, as is the case with sea water, existing in solution as bicarbonate and a proton (or H3O+)…" - Goal of this code is to capture students' responses that discuss the role of pK_a values in determining the direction of equilibria - Does not have to explicit about the values; can simply state relative values ("the pK_a value of H_2CO_3 is lower than that of water"). - **If pK_a values used in an unexpected fashion (as an absolute measure of acidity, rather than relative), stated incorrectly, or mistaken for pH, code as Y**

The formulae in Excel will identify the mode of reasoning based on the presence/absence of combinations of features identified above.

Mode of reasoning	Description	Examples
Descriptive	Describing features of the graph or equations that are visibly shown (*e.g.*, as atmospheric CO_2 goes up, pH goes down).	I disagree with Sam because, based on the equations, H_3O^+ is generated. OR I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down.
Relational	Identifying that increased CO_2 (either atmospheric or aqueous) levels are correlated with H^+ being generated in the equations AND/OR Identifying that increased H^+ being generated in the equations is correlated with decreasing seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. As is shown in the equations, increasing CO_2 produces more H_3O^+ ions. AND/OR I disagree with Sam because, based on the equations, H_3O^+ is generated. Increasing H_3O^+ makes the seawater more acidic, so pH goes down.
Linear causal	Uses Le Chatelier or pK_a values to justify why atmospheric CO_2 influences the direction of the chemical equilibria and connects back to the impact that has on seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. Due to Le Chatelier's principle, increasing CO_2 produces carbonic acid, which leads to increased concentration of H_3O^+ ions. This results in increased acidification of the seawater, and thus a lower pH.
Multicomponent causal	Uses both Le Chatelier and pK_a values to walk through how and why atmospheric CO_2 is influences the direction of the chemical equilibria and connects back to the impact that has on seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. Due to Le Chatelier's principle, increasing CO_2 produces carbonic acid, which leads to increased concentration of H_3O^+ ions. This results in increased acidification of the seawater, and thus a lower pH. Additionally, the pK_a value for carbonic acid is 6.35. Given that the pH of water is around 8, this suggests that Equilibrium 2 favours the products. This means formation of hydronium ions is favoured in Equilibrium 2, which results in a decline in seawater pH.

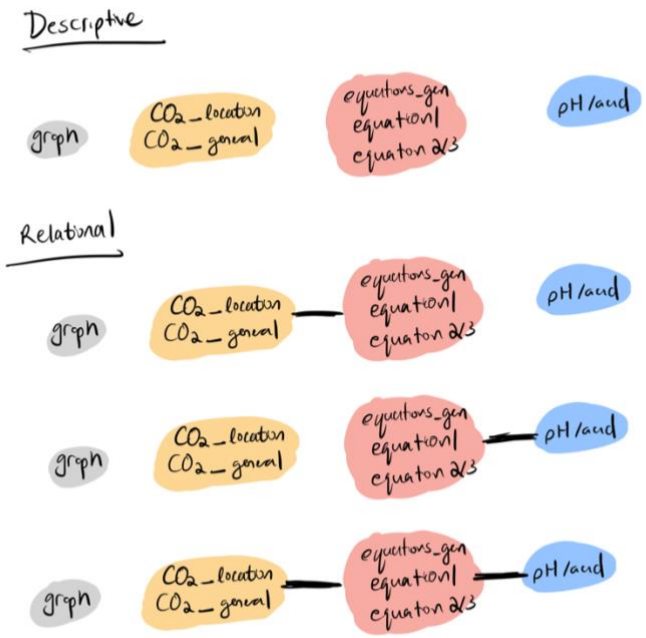
Example of multi-component causal argument:

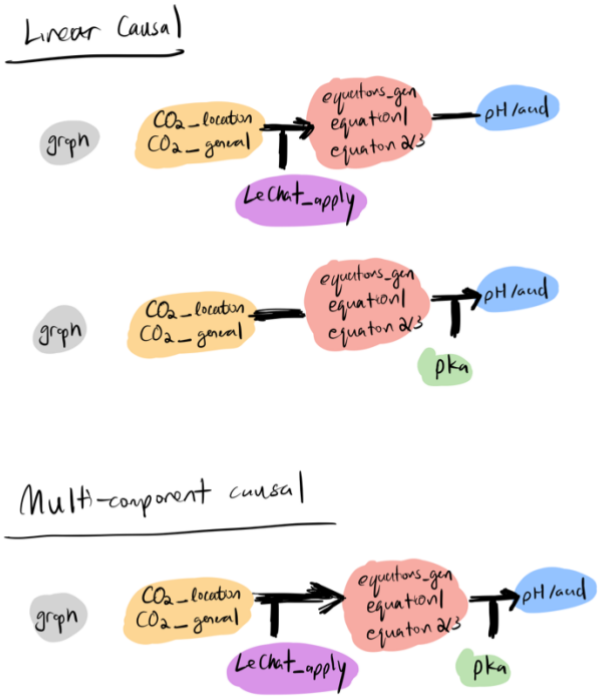
I disagree with Sam. When looking at the graph provided, the atmospheric CO_2 concentration increases as seawater pH decreases. This shows that there is a correlation between atmospheric CO_2 and seawater pH. The chemical equations provided help explain this relationship. As atmospheric CO_2 levels increase, more CO_2 diffuses into seawater. This increase in CO_2 pushes the first equilibrium to form carbonic acid, as Le Chatelier's principle states that

an increase in concentration in reactants will drive an equilibrium to products. Carbonic acid reacts further with water in the second and third equilibria to generate bicarbonate ions and hydronium ions. Increased concentration of hydronium ions increases the overall acidity of the seawater, resulting in a decline in the seawater pH. Therefore, increasing CO_2 levels have a direct impact on the pH of seawater.

Additionally, the pK_a value for carbonic acid is 6.35. Given that the pH of water is around 8, this suggests that second equilibrium favours the products. This means formation of hydronium ions is favoured in the second equilibrium, which results in a decline in seawater pH.

Reasoning diagrams for each mode of reasoning





Inter-rater reliability of qualitative coding

An initial coding protocol was developed by co-author JD using the entire dataset (*N* = 166). Inter-rater reliability (IRR) was measured using Krippendorf's alpha (Krippendorff, 1970) to mitigate inflation of agreement due to chance agreements between raters. The IRR process began with co-authors JD and MR independently coding 15% (*n* = 26) of the full data set with the initial codebook. After this initial round, Krippendorf's alpha was calculated to be 0.372—less than the established threshold for adequate agreement using Krippendorf's alpha (0.67). Discussions between JD and MR following this first round revealed that many of the disagreements between the two raters were consistent—many of the disagreements centred on differing interpretations of how to apply the codes related to participants' descriptions of increasing CO_2 levels. For example, certain arguments would describe CO_2 in a general fashion ("as CO_2 increases, equation 1 shows…"), while other arguments would be more specific ("as CO_2 increases in the atmosphere, it will increase in the seawater…").

After modifications were made to the codebook, a second round of inter-rater coding was conducted. Authors JD and MR independently coded another 15% of the full data set with the modified codebook. After this second round, Krippendorf's alpha was calculated to be 0.908, indicating strong agreement between the two raters. After this second round, JD and MR met a final time to any differences in coding from the second round of IRR, which were often disagreements in whether code was being implied or not within a participants' argument. Following this second, final round of inter-rater reliability, author JD coded the entire data set (*N* = 166) using the finalized codebook.

Participants' arguments were not organized in terms of Eng+ and Eng1st groups during the initial qualitative coding process. This was done to limit bias, ensuring that coders were blind to which arguments were constructed by Eng+ and Eng1st participants.

Generating language profiles from LEAP-Q data

We summarized the data gathered from the LEAP-Q into descriptive English language proficiency and history indices for each respondent (Krizman et al., 2014; Marian, Chabal, Bartolotti, Bradley, & Hernandez, 2014; Reichle & Birdsong, 2014). A single descriptive index for English language proficiency and history was also generated by averaging the individual proficiency and history indices. Distributions for Eng+ and Eng1st respondents are shown in Figure S1.

Statistical analyses

Fischer's exact test

A Fischer's exact test in SPSS was used to test independence of the modes of reasoning between Eng+ and Eng1st responses. Post-hoc pairwise z-tests with Bonferroni adjusted *p*-values for multiple comparisons were used to identify precisely which modes of reasoning differed significantly from each other between Eng+ and Eng1st responses.

Cluster analysis

Cluster analysis was conducted in SPSS using two cluster variables: relative English language history and relative English language proficiency. Hierarchical agglomerative cluster analysis was conducted using Squared Euclidian distance with Ward's linkage method. This method yielded a two and four-cluster solution. The four-cluster solution was used in a subsequent K-means cluster analysis as this solution meaningful insights into participants' language backgrounds by differentiating participants into four different combinations of language proficiency and history.

Logistic regression

Collinearity diagnostics revealed no collinearity between parameters (Table S2). In SPSS, full factorial multinomial logistic regression was used to evaluate the association between English language proficiency and history (independent covariates) and mode of reasoning (dependent categorical output). Academic year was included as a control variable (categorial predictor).

Upon determining that English language history was a significant predictor and reviewing follow-up interview data, an additional full factorial multinomial logistic regression model was generated using time in English country, time in English family, and time in English school/work (independent covariates) and mode of reasoning (dependent categorical output).

Follow-up interviews

Follow-up interviews were conducted with a subset of participants (*n* = 15). Participants were a mix of Eng+ and Eng1st, but all participants self-identified as bi/multilingual during the interviews.

Interviews were conducted using a standardized, open-ended approach (Patton, 2015).We organized responses in terms of question topic and then identified prominent themes and experiences inductively. We paid close attention to two aspects of participants' responses during our analysis: (a) how participants perceived the role of language in *their* own learning experiences and (b) how they perceived the role of language in *others'* learning experiences (peers, friends, classmates). Multiple readings of the transcripts, memoing, and writing allowed us to actively reflect on these aspects throughout analysis.

To establish trustworthiness, member-checking was conducted with participants to gain feedback on analysis or ask follow-up questions to learn more about their perspectives. We also actively searched for disconfirming evidence in the transcripts to ensure that our analysis was not simply conforming to our biases. This process also allowed us to capture the full spectrum of experiences and perspectives in the interview sample.

Follow-up Interviews Protocol

1. Interview Checklist
    - Zoom session scheduled
    - Recording enabled on Zoom
    - Participants' instrument responses retrieved
    - Notebook for notetaking

2. Confidentiality Discussion (5 minutes)

    - Thank participant for their participation
        - I want to begin by thanking you for participating today. Your participation in this interview is very important to me, and I want you to know that your time and effort are very much appreciated.
    - Remind participants that they are allowed to leave the interview at any time.
        - This discussion will take approximately 30 minutes. I understand if you need to leaver sooner.
    - Reminder of use of data and confidentiality
        - Anything you describe to me verbally today will be used to guide our interpretation of your responses. Your identity will be removed from all pieces of data and your name will be changed to a pseudonym for the sake of confidentiality in any future communications.
    - Explanation of use of recording and confidentiality
        - I would like to record this conversation today. The recording is to accurately capture our conversations here today and will be used only for myself and my research group to re-visit if we need to clarify any parts of this discussion.
    - Turn on recording.
    - Any further questions?
        - Do you have any further questions about the discussion itself or the confidentiality of this conversation? If you have any further questions, you may contact either my research advisor (Dr. Flynn) or me through the e-mail addresses I will present in a later slide.

    - Goals for Interview: (1) probe how and why respondents constructed their ocean acidification argument on the argumentation/LEAP-Q instrument; (2) learn about participants' language experiences in chemistry education

3. Introduction (3 minutes)

    - What is the project about?
        - I am collecting data on participants' abilities to use evidence when making decisions about climate change issues. You completed a survey earlier in the year, and we are following up with some participants to learn more about their responses and to hear your interpretation of our conclusions.
    - What am I asking you to do?

- Today I'll simply be asking you a few questions about your response to one of the questions on the survey: constructing an argument in response to a social media about ocean acidification. The questions will focus on how and why you decided to respond in the way that you did.
- Do you have any questions before we begin?

4. Questions about response process (15 minutes)

- Initial approach to item (open-ended)
  - Show respondent ocean acidification item (without their response)
    - How did you approach this question?
    - If you were asked to respond to the question now, how would you respond?

  - Show respondent their response
    - Seeing your response from a few months ago, did you approach the question any differently compared to what you just described to me? How so?
    - Can you recall for me what you were thinking about when you came up with your response?
- Member-checking: We evaluated responses using these four categories. Descriptive means that the response simply described what was available in the question. Relational means that the response make an initial connection between the data. Causal means that the response used chemistry concepts to justify the connection.
    - Your response was said to be _____. Do you agree with this?
    - What do you think a _____ argument looks like?
    - If they produced causal, ask them about descriptive and relational
    - If they produced descriptive or relational, ask them about causal.
    - If there is time at end: show them an example of each and ask them to categorize.

- Verbal probing
  - Ability:
    - Was coming up with your answer easy or hard?
    - You said it was a bit hard. What made it hard?
    - Do you feel like you knew enough about the topic to respond appropriately?
    - Do you remember how long did it took for you to respond?

  - Expectations and stakes:
    - Your answer was longer/shorter than others who responded. Can you share why you decided on this specific level of detail to provide in your response?
    - Could you have provided more detail? If so, what would you have added?

        - Why did you decide not to discuss the equations, Le Chatelier's principle, etc. in your response?
        - Would you change anything in your response if you had more time?
        - Would you change anything in your response if this was on a final exam?
- Language
    - We asked you a question about your first language before the main questions. (As someone who said they are Eng+), to what extent did that influence how you responded, if at all?
    - In general, do you think do better on open-response items or multiple-choice items? Why?
    - Do you think it would have been easier/harder to respond if there were multiple choice options? Why?
    - Was it easy/hard coming up with the words for your answer? Why do you think that?
    - You said you felt _____ when responding in your first language. Can you explain why you feel that way?
    - If this was a test, which language would you prefer to respond in?
    - Do you have any suggestions for how teachers could make changes to support Eng+ students?

5. Wrap-up
    - Is there anything else you would like to add or comment on regarding your experience today, during the survey, or the project in general?
    - Thank you again for participating in this conversation today. I really appreciate it.

6. Additional prompts
    - If participants are silent following initial prompts:
        - Wait. Let people think for a few minutes before jumping in and encouraging a response.
        - If participants still do not respond after a few minutes of silence, provide a starting point for a discussion. This may include:
            - Phrasing the question in a different way
            - Providing an example that participants can build off of and begin discussion
    - Other prompts
        - Please tell me more about that.
        - What do you mean by that?
        - Could you give me an example?
        - I didn't understand that. Can you explain?
        - I hadn't thought of that perspective.
        - Could you go back to…?
        - Could you run that by me again? I'm afraid I still don't understand.
        - What happened that made you think that?

Supplemental Text

Argumentation in education

Arguments and explanations are distinct. An explanation is used to explain an agreed-upon fact or phenomenon (National Research Council, 2012; Osborne & Patterson, 2011), while arguments justify a fact or phenomenon that is not agreed-upon (Kuhn, 2011; McNeill, Lizotte, Krajcik, & Marx, 2006); rather, the claim is in doubt and must be advanced through reasoning by constructing an argument about the fit between the evidence and claim (Osborne & Patterson, 2011; Toulmin, 1958). Arguments therefore provide a an opportunity to investigate how students are reasoning about phenomena (Berland & Reiser, 2009; Emig, 1977; Grimberg & Hand, 2009).

Reasoning has been analysed through a variety of different lenses and frameworks in science education research. These approaches include Type I and II reasoning (Maeyer & Talanquer, 2013; McClary & Talanquer, 2011; Talanquer, 2007, 2017), teleological reasoning (Abrams & Southerland, 2010; Ira Caspari, Weinrich, Sevian, & Graulich, 2018; DeCocq & Bhattacharyya, 2019; Talanquer, 2007; Trommler et al., 2018), abstractedness and abstraction (Sevian, Bernholt, Szteinberg, & Auguste, 2015; Melissa L. Weinrich & Sevian, 2017), rules-, case-, and model-based reasoning (DeCocq & Bhattacharyya, 2019; Kraft, Strickland, & Bhattacharyya, 2010; Windschitl, Thompson, & Braaten, 2008), and causal, mechanistic, and causal mechanistic reasoning (Cooper, Kouyoumdjian, & Underwood, 2016; Crandell, Kouyoumdjian, Underwood, & Cooper, 2018). In this study, we analysed participants' arguments in terms of four modes of reasoning: descriptive, relational, linear causal, and multi-component causal (I. Caspari, Kranz, & Graulich, 2018b; Sevian & Talanquer, 2014; M. L. Weinrich & Talanquer, 2016).

Increasingly, research investigations have explored students' reasoning through their explanations and arguments about scientific phenomena on assessments (I. Caspari et al., 2018a; Ira Caspari et al., 2018; Deng & Flynn, 2019, 2021; A. Moon et al., 2016; Alena Moon et al., 2019; Sevian & Talanquer, 2014; Stowe & Cooper, 2019). Findings from this work suggest that how students' reason and argue varies between tasks and contexts, and for students to learn how to justify claims with sound evidence and reasoning, these practices must be reinforced in various aspects of a course or curriculum (learning outcomes, formative assessments, instructor modeling, *etc*.). From this work, researchers have begun to formulate ways to support students' abilities to construct arguments on assessments in both classroom and laboratory contexts (I. Caspari & Graulich, 2019; Graulich & Caspari, 2021; Petritis, Kelley, & Talanquer, 2020).

English learners in STEM

Language plays an active role in scientific learning; language is a means of doing science and constructing scientific understanding (Abedi, 2002; Afitska & Heaton, 2019; National Academies of Science Engineering and Medicine, 2018; Solano-Flores & Trumbull, 2003; Wolf et al., 2008; Yore et al., 2004). Contemporary approaches have argued that language and content are inseparable (Grapin, Llosa, Haas, & Lee, 2020; Lee, 2021; Lee, Quinn, & Valdés, 2013; Lee & Stephens, 2020; National Academies of Science Engineering and Medicine, 2018), and that by supporting students' abilities to make sense of phenomena in classrooms, students may also

gain language expertise in an environment where a focus on language is explicit. Students view language as a tool to construct meaning about scientific phenomena, rather than something separate from the content (*i.e.*, what language does vs. what language is) (Grapin, 2021; Lee & Stephens, 2020). Through this lens, language is a product of interaction and learning, not a precursor or prerequisite (National Academies of Science Engineering and Medicine, 2018).

At the core of challenges associated with supporting Eng+ learners with instructional accommodations, interventions, resources, *etc*. is determining who qualifies as an Eng+ learner. Eng+ classification has been cited as an issue at various levels of education. there is no common approach to classification and reclassification of Eng+s. For example, in the US, the criteria for both entering Eng+ status and exiting Eng+ status can vary considerably between contexts (National Academies of Science Engineering and Medicine, 2018). Variable classification systems may contribute to an overestimation of achievement gaps in STEM subjects between Eng+s and non-Eng+s. For example, exclusion of recently English-proficient ELs leads to overestimation of achievement gaps in STEM subjects between ELs and non-ELs and underestimation of ELs' STEM proficiency (National Academies of Science Engineering and Medicine, 2018).

Several accommodations and resources to support Eng+ students have been reported. For example, providing extra time, linguistic modifications that simplify linguistic complexity in assessment items, providing bilingual glossaries/dictionaries. Institutional supports such as social-belonging interventions have also been shown to be helpful for strengthening feelings of inclusion for English newcomers (LaCosse, Canning, Bowman, Murphy, & Logel, 2020). However, many classroom-based existing interventions lack evidence to support their proposed impacts and may actually limit Eng+ students' success if applied with careful consideration of students' linguistic abilities and needs (Abedi, 2002; Abedi, Zhang, Rowe, & Lee, 2020; Buxton et al., 2014). Often, the challenges of successfully applying accommodations have been complicated by constrained views of who qualifies as an Eng+ student (Lee, 2021; National Academies of Science Engineering and Medicine, 2018; Wolf et al., 2008). As demonstrated by the interview findings in this study, binary labels such as Eng+ and Eng1st can risk masking the diversity linguistic and cultural backgrounds within both groups. Future research on interventions might adopt some of the methods used here (LEAP-Q, cluster analysis) to generate a more comprehensive understanding of the language proficiency/history within a particular sample to then use this data to identify how interventions impact specific subgroups. Students in this work suggested the following educational supports: course tutorials led by teaching assistants that share a non-English language with students to gain discipline-specific language experience, conducting brief surveys of languages in a classroom to determine what types of linguistic support are needed (if at all), community spaces for Eng+ students (specifically English newcomers) to foster feelings of inclusion and gain language and cultural expertise with their peers. However, perceptions of usefulness and effectiveness are not equivalent to actual effectiveness. If supports are introduced, they should be evaluated for their efficacy and quality.

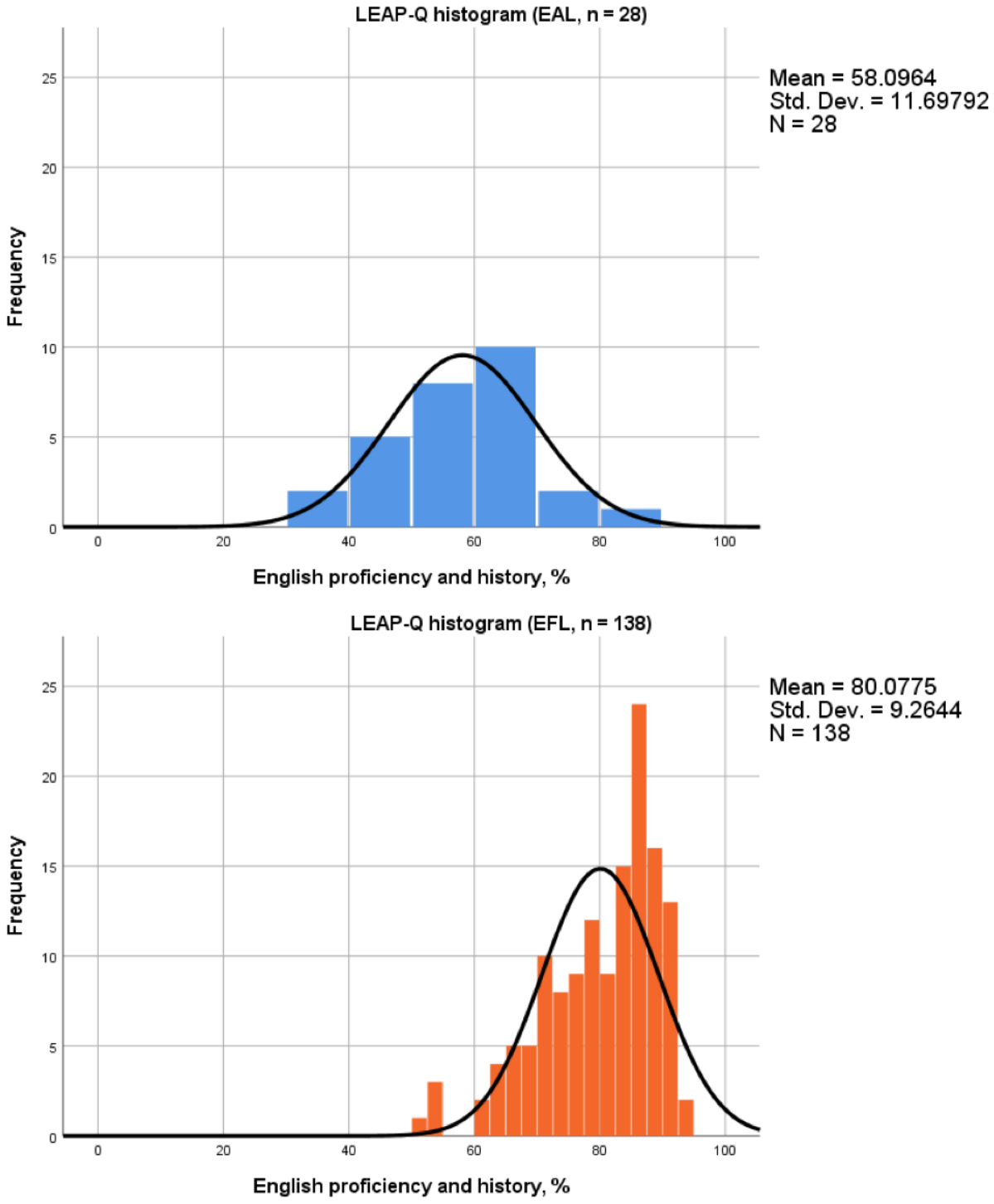


Fig. S1. Distributions of English proficiency and history for Eng+ participants (n = 28, top) and Eng1st participants (n = 138, bottom).

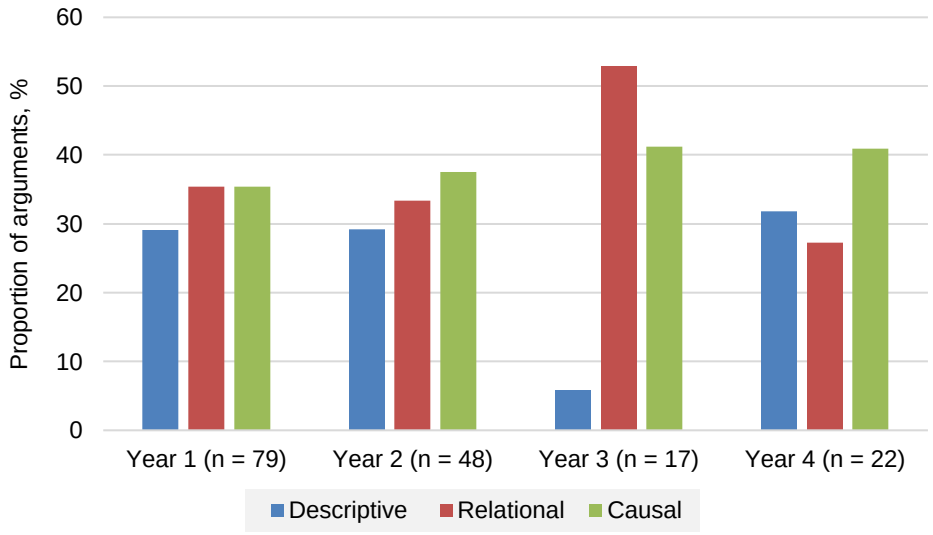


Fig. S2. Reasoning demonstrated in responses per academic year.

Table S1. Descriptions of modes of reasoning and exemplars from dataset

Mode of reasoning	Description	Examples
Descriptive	Describing features of the graph or equations that are visibly shown (*e.g.*, as atmospheric CO_2 goes up, pH goes down).	I disagree with Sam because, based on the equations, H_3O^+ is generated. I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down.
Relational	Identifying that increased CO_2 (either atmospheric or aqueous) levels are correlated with H^+ being generated in the equations Identifying that increased H^+ being generated in the equations is correlated with decreasing seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. As is shown in the equations, increasing CO_2 produces more H_3O^+ ions. I disagree with Sam because, based on the equations, H_3O^+ is generated. Increasing H_3O^+ makes the seawater more acidic, so pH goes down.
Linear causal	Uses Le Chatelier or pK_a values to justify why atmospheric CO_2 influences the direction of the chemical equilibria and connects back to the impact that has on seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. Due to Le Chatelier's principle, increasing CO_2 produces carbonic acid, which leads to increased concentration of H_3O^+ ions. This results in increased acidification of the seawater, and thus a lower pH.
Multicomponent causal	Uses both Le Chatelier and pK_a values to walk through how and why atmospheric CO_2 is influences the direction of the chemical equilibria and connects back to the impact that has on seawater pH.	I disagree with Sam because, based on the graph, CO_2 goes up as pH goes down. Due to Le Chatelier's principle, increasing CO_2 produces carbonic acid, which leads to increased concentration of H_3O^+ ions. This results in increased acidification of the seawater, and thus a lower pH. Additionally, the pK_a value for carbonic acid is 6.35. Given that the pH of water is around 8, this suggests that Equilibrium 2 favours the products. This means formation of hydronium ions is favoured in Equilibrium 2, which results in a decline in seawater pH.

Table S2. Collinearity statistics for logistic regression with English proficiency and history (predictors) and reasoning.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics
		B	Std. Error	Beta	t	Sig.	Tolerance
1	(Constant)	.947	.381		2.488	.014	
	AvgHist	.019	.006	.298	3.110	.002	.620
	AvgProf	-.002	.005	-.046	-.483	.630	.620

Coefficients[a]

Model		Collinearity Statistics
		VIF
1	(Constant)	
	AvgHist	1.613
	AvgProf	1.613

a. Dependent Variable: Reasoning#

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	AvgHist	AvgProf
1	1	2.969	1.000	.00	.00	.00
	2	.018	12.725	.93	.06	.36
	3	.013	15.400	.06	.93	.64

a. Dependent Variable: Reasoning#

Table S3. Logistic regression output. Dependent variable: mode of reasoning. Independent predictors: English history, English proficiency, academic years 1-4.

Reasoning[a]		B (SE)	Sig., *p*	95% Confidence Interval for Exp(B)		
				Lower Bound	Exp(B)	Upper Bound
Descriptive	Intercept	0.810 (1.34)	0.546			
	English history	-0.018 (0.02)	0.343	0.946	0.982	1.019
	English proficiency	0.008 (0.02)	0.626	0.976	1.008	1.040
	[Year=1.00]	-0.407 (0.64)	0.525	0.190	0.666	2.336
	[Year=2.00]	-0.318 (0.67)	0.636	0.195	0.727	2.717
	[Year=3.00]	-2.426 (1.20)	0.044	0.008	0.088	0.935
	[Year=4.00]	0[b]				
Causal	Intercept	-3.532 (1.60)	0.027			
	English history	0.055 (0.02)	0.014	1.011	1.057	1.104
	English proficiency	-0.004 (0.02)	0.779	0.966	0.996	1.026
	[Year=1.00]	-0.078 (0.62)	0.900	0.276	0.925	3.096
	[Year=2.00]	-0.108 (0.65)	0.868	0.253	0.898	3.186
	[Year=3.00]	-0.498 (0.75)	0.508	0.139	0.608	2.660
	[Year=4.00]	0[b]				

a. The reference category is: Relational.

b. This parameter is set to zero because it is redundant.

R^2 = 0.12 (Cox & Snell), 0.14 (Nagelkerke), Model χ^2 (10) = 21.81, $p < .05$

Table S4. Logistic regression output. Dependent variable: mode of reasoning. Independent predictors: Time in English language school, time in English language home, and time in English language country.

Reasoning[a]		B (SE)	Sig., *p*	95% Confidence Interval for Exp(B)		
				Lower Bound	Exp(B)	Upper Bound
Descriptive	Intercept	-0.278 (0.70)	0.690			
	School	0.038 (0.04)	0.322	0.964	1.038	1.119
	Home	-0.014 (0.03)	0.638	0.932	0.986	1.044
	Country	-0.019 (0.05)	0.690	0.896	0.982	1.075
Causal	Intercept	-1.387 (0.80)	0.083			
	School	0.023 (0.03)	0.489	0.958	1.024	1.093
	Home	0.082 (0.04)	0.030	1.008	1.085	1.168
	Country	-0.016 (0.05)	0.765	0.889	0.985	1.090

a. The reference category is: Relational.

R^2 = 0.08 (Cox & Snell), 0.09 (Nagelkerke), Model χ^2 (6) = 13.08, $p < .05$

Table S5. Summary of follow-up interview participants (*n* = 16). Participants had the option to select their own pseudonym and participants who did not have a preference were assigned a pseudonym by the researchers. Pseudonyms assigned by researchers were selected to ensure anonymity while also considering the identities expressed by interviewees during interviews (*e.g.*, cultural/language background).

Pseudonym	Self-identified language group	Primary language of chemistry instruction	Languages (in order of dominance)	Academic year	Reasoning
Elaine	Eng+	French	French, English, Spanish	1st	Relational
Kyle	Eng1st	French	English, French	3rd	Causal
Rosa	Eng+	English	English, French, Spanish	2nd	Descriptive
Melanie	Eng+	French	French, English	2nd	Descriptive
Mohamed	Eng+	French	English, French, Arabic	2nd	Descriptive
Nicki	Eng1st	English	English	2nd	Relational
Steph	Eng+	English	English, Punjabi, Urdu	2nd	Causal
Jessie	Eng1st	English	English, French	2nd	Causal
Serena	Eng1st	French	English, French	1st	Causal
Madeline	Eng1st	French	English, French	2nd	Relational
Jen	Eng1st	English	French, English	4th	Descriptive
Nicholas	Eng+	English	English, Tagalog	3rd	Relational
Jeff	Eng1st	English	English, Spanish, French	2nd	Causal
Adelaide	Eng1st	English	English	2nd	Relational
Peter	Eng+	English	French, English, Kinyarwanda, Spanish	4th	Causal

Table S6. Participants' perspectives on supporting students from non-dominant language backgrounds, including suggested resources and/or ways to support.

How important is it, if at all, to support students from non-English/non-French language backgrounds?	
Student's responsibility (n = 2)	**Kyle**: There's always a possibility of some sort of language barrier. In any university course, I'd say you're kinda on your own. You know, if you take a course in English and you don't know English then that's kind of your fault. It's the harsh reality.
Important, but can be limiting (*n* = 1)	**Mohamed**: Maybe at first it would help because it would allow people time to get comfortable with the language of the country, with the place, the people, the expressions. But in the long term of things, I think it would be bad because you're going to work as a doctor in the future, so you're not going to just work in your first language.
Important; students are underestimated (*n* = 12)	**Rosa**: The resources that we have, like the tutoring centre, the writing workshop centre—all of them are offered in English. If a person speaks Cantonese, there's no Cantonese tutor available that the prof can suggest. So if the professor or the university could provide that, that would be helpful... They're [Eng+ students] not getting special treatment, they're just getting a leg up so they can be equal to everyone else.

Table S7. Eng1st and Eng+ participants' perceptions on how context influences how they respond.

How do you think your answer would have changed in an exam context?	
Would add more detail (*n* = 4)	**Elaine (Eng+)**: I definitely would have elaborated more... I know like increasing the CO_2, it's going to forward the reaction, that's going to increase the H_2CO_3, and then it's going to also increase the H^+ concentration. So if the CO_2 dissolves, the concentrations are going change. I would have added more like something like that on a test.
Would tailor to course expectations (*n* = 8)	**Adelaide (Eng1st):** I feel like I would have been more prepared for it. Or at least like, like going into the survey, I know it's chemistry, but I didn't really know anything specific. But say for a test, I would have reviewed this type of material. So I would have had the vocabulary. And I would have just done a more thorough job and like cited more things from the question.
No change (*n* = 1)	**Steph (Eng+):** I think it'd be the same. Yeah, cause I'm pretty concise with my words. So I think it'd be like, pretty straightforward for me. Because I'm pretty concise writer, my answers are usually pretty short. So I just, I think I would just hit the points, like needed in both situations.
No change, desire to be understood (*n* = 1)	**Peter (Eng+):** I feel like just off the bat, I kind of have this knee jerk reaction that maybe what I'm conveying isn't always clear. So I try to either reinforce my arguments with a lot of like, facts or proofs based on what is already presented. So if I see that we have all the equilibria I'm like, "Okay, how can I use this, to showcase my points in a way that's understandable?

Table S8. Participants' perceptions on how their language backgrounds contributed to how they constructed written arguments in chemistry.

How would your answer change, if at all, if you were asked to respond in [non-English first language]?	
Challenging because learned chemistry in English (*n* = 5)	**Steph (Eng+)**: I would definitely use a lot of English words, just because I wouldn't know the words for the things I want to talk about in Urdu or Punjabi, or they would just be the same [regardless], like acid is still acid. Right? So you would definitely hear like a lot of English...Like, the Urdu and Punjabi I know is more just conversational. Not really scientific.
No change, but takes longer because of translation (*n* = 2)	**Jen (Eng1st)**: In chemistry, my brain actually thinks in English first. I think French might be a bit harder, because I'd have to translate some of the words. Like, I could still do it, but I would have to mentally translate.
Harder in English; learned chemistry in French (*n* = 3)	**Melanie (Eng+):** I've done my whole schooling in French, so all the terms and stuff are in French in my head... In French, instead of saying, "I would disagree with the Twitter user, it is not misleading", I would explain why it was not misleading, instead of just saying it. Because in English, I feel like I affirm more things instead of explaining why it's not as misleading. So I would have probably went into the why a little bit more.
No preference (*n* = 3)	**Kyle (Eng1st):** If I responded to this in French, it would be about the same. I grew up. So I went to a French High School, full French High School. I'm taking all my studies in French now. So I actually know all the concepts in French. So what I do in my head is I I translate that and put it down in English.

Table S9. Output from K-means cluster analysis using four-cluster solution.

Initial Cluster Centers

	Cluster			
	1	2	3	4
Average English history	90.21	54.96	78.69	27.70
Average English proficiency	98.75	86.25	55.00	45.00

Iteration History[a]

	Change in Cluster Centers			
Iteration	1	2	3	4
1	10.171	7.847	11.845	14.980
2	0.674	0.141	3.140	6.867
3	0.341	0.904	1.310	1.592
4	0.000	0.726	0.000	0.808
5	0.000	0.000	0.000	0.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 5. The minimum distance between initial centers is 37.401.

Table S10: Summary of interview themes and quotes (see attached file).

References

Abedi, J. (2002). Assessment and accommodations of English language: Issues, concerns, and recommendations. *Journal of School Improvement*, *3*(1), 83–89.

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining Effectiveness and Validity of Accommodations for English Language Learners in Mathematics: An Evidence-Based Computer Accommodation Decision System. *Educational Measurement: Issues and Practice*, *39*(4), 41–52. https://doi.org/10.1111/emip.12328

Abrams, E., & Southerland, S. (2010). The how's and why's of biological change : How learners neglect physical mechanisms in their search for meaning. *International Journal of Science Education*, *23*(12), 1271–1281. https://doi.org/10.1080/09500690110038558

Afitska, O., & Heaton, T. J. (2019). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, 1396–1422. https://doi.org/10.1002/sce.21545

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Barbera, J., & VandenPlas, J. R. (2011). All assessment materials are not created equal: The myths about instrument development, validity, and reliability. *ACS Symposium Series*, *1074*, 177–193. https://doi.org/10.1021/bk-2011-1074.ch011

Berland, L. K., & Reiser, B. J. (2009). Making Sense of Argumentation and Explanation. *Science Education*, *93*, 26–55. https://doi.org/10.1002/sce.20286

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Buxton, C., Allexsaht-Snider, M., Aghasaleh, R., Kayumova, S., Kim, S.-H., Choi, Y.-J., & Cohen, A. (2014). Potential Benefits of Bilingual Constructed Response Science Assessments for Understanding Bilingual Learners' Emergent Use of Language of Scientific Investigation Practices. *Double Helix*, *2*(September 2016), 1–21.

Caspari, I., & Graulich, N. (2019). Scaffolding the structure of organic chemistry students' multivariate comparative mechanistic reasoning. *International Journal of Physics and Chemistry Education*, (July). https://doi.org/10.13140/RG.2.2.35136.48646

Caspari, I., Kranz, D., & Graulich, N. (2018a). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chemistry Education Research and Practice*. https://doi.org/10.1039/C8RP00131F

Caspari, I., Kranz, D., & Graulich, N. (2018b). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chemistry Education Research and Practice*, *19*(4), 1117–1141. https://doi.org/10.1039/C8RP00131F

Caspari, Ira, Weinrich, M. L., Sevian, H., & Graulich, N. (2018). This mechanistic step is "productive": organic chemistry students' backward-oriented reasoning. *Chemistry Education Research and Practice*, *19*(1), 42–59. https://doi.org/10.1039/c7rp00124j

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating Students' Reasoning about Acid-Base Reactions. *Journal of Chemical Education*, *93*(10), 1703–1712.

https://doi.org/10.1021/acs.jchemed.6b00417
Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2018). Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. https://doi.org/10.1021/acs.jchemed.8b00784
DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: a literature review. *Quality and Quantity*, *52*(4), 1523–1559. https://doi.org/10.1007/s11135-017-0533-4
DeCocq, V., & Bhattacharyya, G. (2019). TMI (Too much information)! Effects of given information on organic chemistry students' approaches to solving mechanism tasks. *Chemistry Education Research and Practice*, *20*(1), 213–228. https://doi.org/10.1039/C8RP00214B
Deng, J. M., & Flynn, A. B. (2019). Characterising students' abilities to argue for "why" when selecting a base to drive an equilibrium to products. *Chemistry Education Research and Practice*, Submitted.
Deng, J. M., & Flynn, A. B. (2021). Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items. *Chemistry Education Research and Practice*, *22*, 749–771. https://doi.org/10.1039/x0xx00000x
Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Masked translation priming effects with low proficient bilinguals. *Memory and Cognition*, *39*(2), 260–275. https://doi.org/10.3758/s13421-010-0004-9
Emig, J. (1977). Writing as a Mode of Learning. *Coll. Compos. Commun.*, *28*(2), 122–128.
Grapin, S. E. (2021). Toward asset-oriented and definitionally clear terminology: A comment on González-Howard and Suárez (2021). *Journal of Research in Science Teaching*, *58*(5), 753–755. https://doi.org/10.1002/tea.21686
Grapin, S. E., Llosa, L., Haas, A., & Lee, O. (2020). Rethinking instructional strategies with English learners in the content areas. *TESOL Journal*, 1–12. https://doi.org/10.1002/tesj.557
Graulich, N., & Caspari, I. (2021). Designing a scaffold for mechanistic reasoning in organic chemistry. *Chemistry Teacher International*, *3*(1), 19–30. https://doi.org/10.1515/cti-2020-0001
Grimberg, B. I., & Hand, B. (2009). Cognitive pathways: Analysis of students' written texts for science understanding. *International Journal of Science Education*, *31*(4), 503–521. https://doi.org/10.1080/09500690701704805
Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x
Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2019). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism*, 8–13. https://doi.org/10.1017/S1366728919000038
Kraft, A., Strickland, A. M., & Bhattacharyya, G. (2010). Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chemistry Education Research and Practice*, *11*(4), 281–292. https://doi.org/10.1039/C0RP90003F
Krippendorff, K. (1970). Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educational and Psychological Measurement*, *30*(1), 61–70.
Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and*

Language, *128*(1), 34–40. https://doi.org/10.1038/jid.2014.371
Kuhn, D. (2011). *The skills of argument*. New York, New York, USA: Cambridge University Press.
LaCosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-belonging intervention improves STEM outcomes for students who speak English as a second language. *Science Advances*, *6*(40), 1–11. https://doi.org/10.1126/sciadv.abb6543
Lee, O. (2021). Asset-oriented framing of science and language learning with multilingual learners, (April), 1–7. https://doi.org/10.1002/tea.21694
Lee, O., Quinn, H., & Valdés, G. (2013). Science and Language for English Language Learners in Relation to Next Generation Science Standards and with Implications for Common Core State Standards for English Language Arts and Mathematics. *Educational Researcher*, *42*(4), 223–233. https://doi.org/10.3102/0013189X13480524
Lee, O., & Stephens, A. (2020). English Learners in STEM Subjects: Contemporary Views on STEM Subjects and Language With English Learners. *Educational Researcher*, *49*(6), 426–432. https://doi.org/10.3102/0013189X20923708
Maeyer, J., & Talanquer, V. (2013). Making Predictions About Chemical Reactivity: Assumptions and Heuristics. *Journal of Research in Science Teaching*, *50*(6), 748–767. https://doi.org/10.1002/tea.21092
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)
Marian, V., Chabal, S., Bartolotti, J., Bradley, K., & Hernandez, A. E. (2014). Differential recruitment of executive control regions during phonological competition in monolinguals and bilinguals. *Brain and Language*, *139*, 108–117. https://doi.org/10.1016/j.bandl.2014.10.005
McClary, L., & Talanquer, V. (2011). Heuristic reasoning in chemistry: making decisions about acid strength. *International Journal of Science Education*, *33*(10), 1433–1454. https://doi.org/10.1080/09500693.2010.528463
McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1
Moon, Alena, Moeller, R., Gere, A. R., & Shultz, G. V. (2019). Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification. *Chemistry Education Research and Practice*, *20*(3), 484–494. https://doi.org/10.1039/c9rp00005d
Moon, A., Stanford, C., Cole, R., & Towns, M. (2016). The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry. *Chemistry Education Research and Practice*, *17*(2), 353–364. https://doi.org/10.1039/c5rp00207a
National Academies of Science Engineering and Medicine. (2018). *English Learners in STEM Subjects: Transforming Classrooms, Schools, and Lives*. Washington, D.C.: The National Academies Press. https://doi.org/10.17226/25182
National Research Council. (2012). *A Framework for K-12 Science Education*. Washington, D.C.: National Academies Press. Retrieved from https://www.nap.edu/read/13165/chapter/1
Osborne, J. F., & Patterson, A. (2011). Scientific Argument and Explanation: A Necessary

Distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438
Patton, M. Q. (2015). Distinguishing Interview Approaches and Types of Interviews. In *Qualitative Research & Evaluation Methods* (pp. 641–648).
Petritis, S. J., Kelley, C., & Talanquer, V. (2020). Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation. *Chemistry Education Research and Practice*, *22*, 105–121. https://doi.org/10.1039/D0RP00268B
Reichle, R. V., & Birdsong, D. (2014). Processing focus structure in L1 and L2 French : L2 proficiency effects on ERPs. *Studies in Second Language Acquisition*, *36*(3), 535–564. https://doi.org/10.1017/S0272263113000594
Sevian, H., Bernholt, S., Szteinberg, G. A., & Auguste, S. (2015). Use of representation mapping to capture abstraction in problem solving in different courses in chemistry. *Chemistry Education Research and Practice*, *16*(3), 429–446. https://doi.org/10.1039/c5rp00030k
Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: a learning progression on chemical thinking. *Chem. Educ. Res. Pract.*, *15*(1), 10–23. https://doi.org/10.1039/c3rp00111c
Solano-Flores, G., & Trumbull, E. (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners. *Educational Researcher*, *32*(2), 3–13. https://doi.org/10.3102/0013189X032002003
Sorana Laneanu, M. (2016). *Response Processes as a Source of Validity Evidence in Self-Report Measures: Theoretical and Methodological Implications*. The University of British Columbia (Vancouver). Retrieved from www.iranesrd.com
Stowe, R. L., & Cooper, M. M. (2019). Arguing from Spectroscopic Evidence. *Journal of Chemical Education*, *96*(10), 2072–2085. https://doi.org/10.1021/acs.jchemed.9b00550
Talanquer, V. (2007). Explanations and Teleology in Chemistry Education. *International Journal of Science Education*, *29*(7), 853–870. https://doi.org/10.1080/09500690601087632
Talanquer, V. (2017). Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, *94*(12), 1805–1810. https://doi.org/10.1021/acs.jchemed.7b00427
Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press. https://doi.org/10.1080/00048405985200191
Trommler, F., Gresch, H., Hammann, M., Trommler, F., Gresch, H., & Hammann, M. (2018). Students' reasons for preferring teleological explanations. *International Journal of Science Education*, *40*(2), 159–187. https://doi.org/10.1080/09500693.2017.1404658
Weinrich, M. L., & Talanquer, V. (2016). Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product. *Chemistry Education Research and Practice*, *17*(2), 394–406. https://doi.org/10.1039/c5rp00208g
Weinrich, M., & Talanquer, V. (2016). Mapping Students' Modes of Reasoning When Thinking About Chemical Reactions Used to Make a Desired Product. *Chem. Educ. Res. Pract.*, *17*(394–406). https://doi.org/10.1039/C5RP00208G
Weinrich, Melissa L., & Sevian, H. (2017). Capturing students' abstraction while solving organic reaction mechanism problems across a semester. *Chemistry Education Research and Practice*, *18*(1), 169–190. https://doi.org/10.1039/c6rp00120c
Willis, G. (2011). Introduction to Cognitive Interviewing. *Cognitive Interviewing*, 2–11. https://doi.org/10.4135/9781412983655.n1
Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the Scientific Method: Model-

Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Science Education*, *92*(5), 941–967. https://doi.org/10.1002/sce.20259

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses (CRESST Report 732). *Review Literature And Arts Of The Americas*, 1–96.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly*, *39*(3), 347–352. https://doi.org/10.1598/RRQ.39.3.8

Zumbo, B. D., & Hubley, A. M. (2017). *Understanding and Investigating Response Processes in Validation Research*. *Springer*. Retrieved from http://link.springer.com/10.1007/978-3-319-56129-5%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/25723322%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4379516%0Ahttp://www.mdpi.com/2072-6651/7/3/648

13.3 Appendix 3: Chapter 6 Supporting Information

Supporting Information: “I am working 24/7, but I can't translate that to you”: The barriers, strategies, and needed supports reported by chemistry trainees from English-as-an-additional language backgrounds

Jacky M. Deng[1], Alison B. Flynn[1,*]

[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie, Ottawa, Ontario, K1N 6N5, Canada.

*E-mail: alison.flynn@uOttawa.ca

TABLE OF CONTENTS

QUALITATIVE CODEBOOK *1*

SUPPLEMENTAL TEXT *3*

THEORETICAL FRAMEWORK: ENGLISH LEARNERS IN SCIENCE, TECHNOLOGY, ENGINEERING, AND MATHEMATICS (STEM) EDUCATION 3

LITERATURE ON POTENTIAL ACCOMMODATIONS AND RESOURCES FOR ENG+ LEARNERS 4

METHODS DETAILS4

Data collection 4

Data analysis 4

REFERENCES 4

DATA FROM PARTICIPANTS' PRE- AND POST-SESSION QUESTIONNAIRE RESPONSES *6*

Table S1: Participants' pseudonyms and first language(s) (*N* = 18). 6

Figure S1: Participants' responses to “In what situations do language barriers arise the most for you? Check all that apply.”6

Figure S2: Participants' ethno-racial backgrounds (*N* = 18). 7

Figure S3: Participants' self-reported proficiencies understanding, writing, speaking, and reading in English (*N* = 18). 7

Figure S4: Participants' career stages, including years (*N* = 18). 8

Figure S5: Participants' responses to: “What types of English language supports/services have you used or currently use? Check all that apply.” 8

QUALITATIVE CODEBOOK

RQ1: What are Eng+ trainees language experiences prior to working and learning in English research programs?		
Theme 1: Eng+ trainees arrive at English institutions with a diversity of language experiences		**Exemplar quotes**
1.1 Limited	Limited opportunities to learn or use English in schooling, home, day-to-day, and/or in scientific contexts prior to enrolling current program.	J.M.D.: When you were doing your undergrad, did you do your education in English? Or was it in your first languages? Participant: Yeah, no English. No, no, at least in Brazil. It's all in Portuguese. Everything in Iran using Persian, even [scientific] references and source books and everything we can find translated into Persian. It can be easy to read these sources and these references in Persian, and because of that, I think everyone prefer to read in Persian.
1.2 Sporadic	Attending sporadic (one-time) courses (in-school, extracurricular, *etc.*) or sporadically learning English to accomplish specific tasks within an otherwise non-English environment (reading books/papers, watching TV, listening to music). Aspects of science education were in English (reading papers, textbooks, *etc.*) but majority conducted in non-English first language.	I've been doing classes, they have these courses outside of school that you go, you watch for one hour English, and you come back and speak Portuguese for the rest of the week. I attend[ed] this course since [I was] 10 years [old]. I don't think it actually helped—it helped having a notion of what the language is, but not going and talking... And, actually, we had an “English for chemistry students” course during my undergrad, but it was just one course for all four years of undergrad.
1.3 Immersive	Consistent opportunities to learn and use English in various social and academic settings (attending an English school, conducting research in an English lab). Consistent opportunities to learn and use English through science education	I went to an international school, so everything I learned was in English. We did have Arabic classes as language classes, but everything we took was in English. Yeah, all my sciences, all the classes I've learned, I've basically learned all my classes in English except for my French class and my Arabic class, which were taught in Arabic and English, and Islamic Studies, which was also taught in Arabic. But everything else is basically English.
Theme 2: Eng+ trainees have varied perspectives on learning and using English in science		**Exemplar quotes**
2.1 Fair to have to learn and use English in science	Argues that it is fair to have to learn English to participate in science. Reasons include: Eng+ students made the deliberate choice to come to an English-speaking institution, reality of the situation, long-term pros of learning English outweigh the short-term cons *etc.*	I mean, it's sad, but I think it's fair. Because we are living in an English-speaking country, and English now is the language of science. So for myself, I think also that it's fair to learn English for science, because science is international, and we have to get the information from all over the world. So it's very important to have just one language, so we understand each other. So it's just, it's just an extra asset to me. To learn a new language. To be obligated to learn a new language that I'll be able to use to communicate with anybody else in the world. It's kind of cool.
2.2 Issue is not learning English, but access to resources/supports to learn	Argues that learning English is fair, but inaccessibility of resources and supports is unfair	What is really unfair is the access to learning English, to English education. That's where the unfairness lies, I think. One thing I did not mention is that the obligation to learn English, in some countries, limits the opportunities to people from wealthy backgrounds. English schools and exchanges are expensive, and a lot of families cannot afford them. Although one universal language is a way to increase

		communication among researchers, it is also a way to exclude people from less economically favored backgrounds.
2.3 Loss of scientific knowledge and expertise	Argues that expecting scientists to communicate their research in English results in loss of scientific knowledge and expertise by limiting opportunities for those who primarily conduct research in languages other than English.	But in terms of like being successful in science without knowing English if we were in our country, that's not fair. Because if you realize a lot of successful scientists or researchers in different countries, they didn't have a lot of opportunity as like a English native speakers because they were they were lacking the opportunity, for example, for networking because they couldn't speak English. They had to do in their own country and that limited a lot of their opportunities. That is not fair. Um, no. I don't think it's fair. In a point that should be able to succeed in science, you have to learn English. And I was reading an article, I think [it] was Nature Science, about the difficulty of researchers from other countries that are not English speakers to have their work recognized specifically because they don't speak English. And this is really challenging because we see really good researchers that don't have English [as a] first language and they don't compete with other researchers that have English. I don't think this is fair, because we're missing a lot on the quality of researchers.
RQ2: What barriers have Eng+ trainees faced while learning, communicating, and doing chemistry in English?		
Theme 1: Learning chemistry is limited by the dominance of English in scientific texts and speech		**Exemplar quotes**
1.1 Struggling to learn chemistry because unfamiliar with English structure, grammar, and/or vocabulary	Struggling to retrieve key messages from text or speech due to challenges associated with English syntax, grammar, and/or vocabulary (non-discipline-specific aspects of scientific English).	When, for example, when I was focusing more on the reading, I was just missing the topic. I was focusing on the words—what's the meaning of, for example, this one and this one. I was reading the whole paragraph, but once the paragraph finished, I didn't understand what's happening with the scientific [content]. And I had to re-read again to understand, 'Okay, what's the concept behind these things?' Yeah, that one was difficult.
1.2 Struggling to learn chemistry because unfamiliar with English scientific jargon	Struggling to retrieve key messages from text due to challenges associated with English scientific jargon. Participant may attribute to challenges with connecting prior expertise (non-English) to content of the text.	And I remember opening the first paper and as I was reading I was like holy sh*t. All these words don't make any sense. Like I speak English and I do understand English, but these words are not like English, English. These are scientific words, just scientific jargon.
1.3 Struggling to understand different accents	Struggling to understand speaker accents during conversations, presentations, discussions, *etc.*	[My labmates] describe it as a bit more like a Southern USA accent. And sometimes he's [labmate] talking and he says something and I'm like, I didn't get it because I don't know that accent… So it makes interaction with him a bit more difficult just because sometimes it's in my head like, 'Oh, we're going to talk? What if I don't understand? How do I react?'
1.4 Struggling to follow speech due to the pace and/or volume of content	Struggling to follow conversations, presentations, discussions, *etc.* due to the pace and/or volume of content. Participant may describe as feeling "overwhelmed" or "overloaded".	Yeah, it's just because sometimes [my supervisor] comes to help me or like to see what I'm doing in the lab. We're doing something, and then he goes through something fast with me and it's like "yeah, you just got to do this and this and this." Some stuff that they think is super easy and you just got to do it. And in my head I'm just trying to process everything he's saying, but he goes through it so fast.
Theme 2: Communicating chemistry is limited by the expectation to write and speak in academic English		**Exemplar quotes**
2.1 Struggling to use English structure, grammar, and/or vocabulary	Struggling to write or speak in English due to challenges associated with English syntax, grammar, and/or vocabulary (non-discipline-specific aspects of scientific English).	Writing has] been a bit hard to be honest. I could definitely write longer texts in Spanish. Both papers—I have two papers—and it was a bit of a hell to write them because all my ideas definitely are simpler when I write in English. And I find it a bit more difficult to like create connections between the sentences. I also tend to write longer sentences because in Spanish, those are allowed and in English, they tend to be more concise. So I tend to overdo it. And sometimes I try to shorten things too much. And then I need to explain more.
2.2 Struggling to communicate in scientific English	Struggling to write or speak in English due to challenges associated with using English in the appropriate scientific register (*i.e.*, speaking formally).	Sometimes I want to make the manuscript look more professional, so I'm trying to use like more sophisticated words that sometimes aren't used in everyday conversational life. So that would take a little bit more energy compared to writing in my own language.
2.3 Concerned with being viewed as less knowledgeable by others	Fear of writing or speaking in English due to concerns of being viewed as less knowledgeable by colleagues when using English. Feelings consistent with impostor syndrome (doubting their skills, talents, or accomplishments; persistent fear of being exposed as a fraud).	But with some native speakers, basically, you're thinking that they see you as the kind of dumb people because you basically struggling with pronunciation and all that stuff. You want for them to have a great opinion about you. Even if you know chemistry, if you kind of struggling with language, it kind of diminishes your value. The fear of sounding dumb or being unable to express ourselves does influence my performance, as one of the students mentioned, is a constant feeling of imposter syndrome, which led me to self-sabotage and demotivation. Also, the fear of saying something wrong limited some networking opportunities since I rather not socializing with them engaging in a conversation.
2.4 Struggling to produce immediate responses	Struggling to effectively participate in scientific discourse when having to improvise or not having preparation prior to generate a response in a limited amount of time with appropriate structure, in academic register, *etc.*	If I don't have a preparation beforehand or if I'm talking as I do the presentation to the people to the person [improvising], I would not be clear. And especially when people ask me questions, answering with something logical—introduction, development and conclusion—it's very hard. So if I don't have a preparation beforehand, the presentation is just not good.
Theme 3: Learning and communication challenges contribute to delays in research		**Exemplar quotes**
3.1 Experiencing research delays due to resource implications	Loss of research time due to more time spent on preparing communication tasks (writing, presentation prep, *etc.*). Struggling to extract information from readings and presentations that are relevant to participants' research topics and goals	It doesn't occur with like experimental work that we are doing. It more hurts when you have to write texts, like way longer time to write, and it kind of slows your pace. I think writing is challenging even in my first language. At least for me, it's my nightmare. Doesn't matter which language I'm speaking. But of course, writing English adds another layer of challenge on writing, specific for scientific writing because it requires a specific vocabulary, requires more grammar. So for me, yes, I do think it's challenging and I have a feeling it takes for me longer than the other students to write. So to write a normal paper, I see students sitting down in one month, they write the paper and I'm trying, it's been six months. I can't [laughs] I just can't. No paragraphs are coming out. So for me, it's what I call a nightmare—writing. And so I've had a hard time with like, writing papers—like it takes me longer than like an English speaking person, because I'm mostly translating in my head and like writing it out as if I was speaking it and saying it.

		I think at the end of the day, the people who know the language better have a better grip on writing, and takes them less time. I think, like the people who do not have it as a first language takes more time. Or sometimes, even I would say that sometimes I use the paraphrasing tool to put my words correctly.
3.2 Experiencing research delays due to social and psychological implications	Avoiding interactions or avoiding asking for help with research from colleagues/supervisors to avoid conflict. Struggling to communicate and establish relationships in their group, making it harder to do research.	So the thing is, all the courses I took were actually really important for my research. They were linked. So because I couldn't understand a lot of the things that were said during the courses, it drained my energy in a way that I felt a little bit... you know, the impostor syndrome? Where you're like 'oh everyone here understands everything that the prof is saying, but no, not me.' So I didn't feel okay. I was emotionally feeling bad. So that impacted my research because I felt like I was not in the right place. It was emotionally draining Because I feel like I can do way more if things were better in the terms of like language, being able to communicate in the lab and, you know, skip some of the steps that I have to do myself [by asking] somebody else [who] can do it with me. But I don't know how to approach them...Like, I don't know how to approach. Like in chemistry, you can just go to a crystallographer and ask for the final thing. Like, I don't have to go through the literature. But I don't know how to approach them!
RQ3: What strategies and supports have Eng+ trainees found to be helpful and/or needed?		
Theme 1: Demonstrating resilience in face of challenges		**Exemplar quotes**
1.1 Independently seeking out resources and support	Evidence of efforts to seek support, adapt to new situations, environments, and communication styles quickly and independently because of having to learn a new language/culture. Including: using translation technologies (e.g., Grammarly, Google Translate, dictionary), seeking extra opportunities reading/listening more, finding friends to share experiences/strategies.	Yeah I mean this is a quite important because when you arrive to a lab, you need to learn a lot of new things. And you can't especially, when you're at a higher level, when you're a postdoc, you cannot rely on other people to teach you things—you're not an undergrad anymore. You also need to be flexible, because you need adapt to the environment you arrive to, and to the people you treat and to the science... So you need a higher degree of flexibility, adaptability to accomplish things. But from like very beginning, I started to go as many seminars that department had just to get used to different accents and different rates of speech. So I had a couple of students say 'Oh it was so great listening to you like, where are you from? I'm from a different country, too. I felt comfortable. I felt so comfortable with you talking that I felt comfortable making question for you.' For me this was really nice, and a lot of people that identified with me when I was presenting. Don't be ashamed. And just try to join in on events as much as possible. It doesn't matter that what is the topic—any conference, any seminars, any events, or just anything that students are trying to gather—just join. And find friends; being alone, it's not a good way.
1.2 Adopting a growth mindset	Demonstrate a growth mindset towards language learning and/or learning in general. Frames challenges as opportunities to improve; unafraid of making mistakes	Working on self is more important than what others tell you. If you try to improve into yourself, because it's a language, you will not get it in a day. And it comes with practice. So if you do the practice on your part, I guess with time it improves. It's not something that difficult that you cannot learn. So I think there should be a will to learn and if the person has that, it can be done easily. It doesn't matter. Yes, if you're in the positive environment, it's encouraging. But if you're not, then you do your own part and improve. We can't change our supervisor, our society, or people, but we can easily change ourselves. We can be positive about that. Don't give up, and try more and more to be able to actually improve our English.
Theme 2: Institutional resources that are discipline-specific and foster community.		**Exemplar quotes**
2.1 Discipline-specific	States need for discipline-specific communication resources (*e.g.* department-based resource), compared to general communication resources (*e.g.*, university academic writing center).	They send us to the English writing center. I found that when the TA or the center don't have a strong science background, then they couldn't give like a lot helpful advice when we get to things that are more academic or more science. So it would be helpful... for example, I'm doing biochemistry, so if there was anyone like have like chemistry background or biology backgrounds that can do the proofreading, then I'm sure she can give more advice on the writing other than the general one.
2.2 Eng+ communities	States need for opportunities that allow Eng+ students to develop English communication skills alongside other students while also using English comfortably with other learners, building social connections, and fostering a sense of belonging.	What helped me was meeting people that were passing through the same thing. And in my case, I was lucky to have students that were joining to group and joining the department that I got to meet and they had more or less same background or going through the same thing. Really early on I was passing through some issues and I felt that was the only one passing through them. And when you listen to someone who is passing the same thing, it's like okay, I'm not that bad. It's just more normal, it will just pass. So I think for me, it helps getting this pure help, this conversation, this communication. And sometimes learning your problem's not that big. You are not stupid, you're not dumb. That is just part of the process. This helps a lot.
Theme 3: Supervisors who are empathetic of Eng+ challenges and prioritize student learning		**Exemplar quotes**
3.1 Patience and empathy	Cites need for supervisors who are exhibit communication styles that demonstrate understanding of Eng+ students' challenges and patience and empathy when interacting with trainees on a day-to-day basis	First thing is they should be understanding. And if they do not understand the situation well, then it's a problem...If you take that into consideration the efforts of the person, and you check back on the concept and not on the phrase that they are writing, I think that would be better on the supervisor's part. First of all, do not expect them to just learn it by night, in one night, and to just be proficient and to have a great relationship with everybody, because not everybody will include them just because they don't speak the same language or they don't speak as well as they would expect. Don't expect them to follow everything that is said and follow your direction properly.
3.2 Providing trainees with resources and supports	Cites need for supervirors who provide tangible support for trainees' development, including supporting attendance of communication courses, providing constructive feedback on writing/presentations aligned with students' goals, fostering an inclusive group atmosphere, offering trainees with flexible deadlines	And I think that both university department and supervisor should encourage a student to take part in these [language] classes instead of put the blame on them, and instead of discouraging them. He's not forgetting that if science doesn't work, that doesn't mean that the student doesn't work. He never forgot that. And that's great. It's really good that someone doesn't forget that that your success is not based solely on your work. It could just be because as a whole, with your PI, the path that has been taken is just not the right one to be successful on. But that's okay, that's what you discover—the research it fails, you take another path, right? Whenever he's [supervisor] giving feedback, you feel that he's giving it with passion, with improvement in mind...Whenever he correcting us, you feel that he's not there to throw you down.

	He's just adding, he always starts with 'Oh, this is so good. The way you write is so good. It's just at the end that we should just add something like this…' So he always has this passion to throw you up instead of down. Like when people see my presentations now they look at them and they're like, "Okay, you've got really nice slides." They're still not perfect up and to my supervisors specific standard, I still haven't gotten to that, but they keep on telling me that I've improved. But it's just, I'm doing a two year master's, so I'm almost done with my first year, and, you know, as much as I'm happy I got better with how my presentation looks and how like the design is, I still haven't got feedback that I wanted.

SUPPLEMENTAL TEXT

THEORETICAL FRAMEWORK: ENGLISH LEARNERS IN SCIENCE, TECHNOLOGY, ENGINEERING, AND MATHEMATICS (STEM) EDUCATION

Language and content are inseparable.[1–5] Students view language as a tool to construct meaning about scientific phenomena, rather than something separate from the content (*i.e.*, what language does vs. what language is).[1,6] Through this lens, language is a product of interaction and learning, not a precursor or prerequisite.[5]

Learning science involves making meaning of scientific knowledge and vocabulary. Language plays an active role in scientific learning; language is a means of doing science and constructing scientific understanding.[5,7–11] Prior knowledge and personal experience play key roles in acquiring new knowledge, as learning and understanding occur when students successfully integrate new information with prior experience and construct new knowledge. Because the knowledge Eng+ students bring to the learning process may differ from Eng1st students, linking what students already know with what they are expected to learn can be a challenge for Eng+ students.

Communicating science describes the ways that ideas are developed, organized, and ultimately communicated in science. Communication patterns vary across languages and cultures and students from diverse language backgrounds often have different interpretations of verbal communication. When communicating in English, Eng+ students may also use the discourse patterns of their first languages. Although alternative communication patterns can provide Eng+ students with powerful ways of demonstrating their knowledge and understanding, these patterns may be ignored or even perceived as disruptive unless teachers recognize and value them.

Doing science involves engaging in inquiry by manipulating materials, making observations, proposing explanations, interpreting and verifying evidence, and constructing ideas to make sense of the world. The emphasis on inquiry has been shown to pose challenges for Eng+ students, as students from oral language traditions may have difficulty using language functions, such as reflecting, predicting, inferencing, and hypothesizing. Newly arrived students may experience difficulties with scientific inquiry in school because they have not been encouraged to ask questions or devise plans for investigation on their own. Students from cultures that respect authority may be receptive to teachers telling and directing them, rather than to inquire, explore, and seek alternative ways.

In this work, aligned with the learning, communicating, doing framework proposed by Lee and Fradd (1998), we investigated how Eng+ chemistry trainees' experiences with learning, communicating, and doing chemistry in English. This framework guided each aspect of our research design, including our research questions (specifically RQ2), interview and focus group protocols, data analysis, and reporting of our results. For example, during the interviews and focus groups, participants were explicitly prompted about their experiences with reading and listening to chemistry content in English (*i.e.*, learning).

LITERATURE ON POTENTIAL ACCOMMODATIONS AND RESOURCES FOR ENG+ LEARNERS

At the core of challenges associated with supporting Eng+ learners with instructional accommodations, interventions, resources, *etc.* is determining who qualifies as an Eng+ learner. Eng+ classification has been cited as an issue at various levels of education as there is no common approach to classification and reclassification of Eng+. For example, in the US, the criteria for both entering Eng+ status and exiting Eng+ status can vary considerably between contexts.[5] Variable classification systems may contribute to an overestimation of achievement gaps in STEM subjects between Eng+ and non-Eng+. For example, exclusion of recently English-proficient ELs leads to overestimation of achievement gaps in STEM subjects between Eng+ and Eng1st and underestimation of Eng+ STEM proficiency.[5]

Several accommodations and resources to support Eng+ students have been reported. For example, providing extra time, linguistic modifications that simplify linguistic complexity in assessment items, providing bilingual glossaries/dictionaries. Institutional supports such as social-belonging interventions have also been shown to be helpful for strengthening feelings of inclusion for English newcomer.[12] However, many classroom-based existing interventions lack evidence to support their proposed impacts and may actually limit Eng+ students' success if applied with careful consideration of students' linguistic abilities and needs.[10,13,14] Often, the challenges of successfully applying accommodations have been complicated by constrained views of who qualifies as an Eng+ student.[3,5,11]

METHODS DETAILS

Data collection

We conducted this study at a bilingual (English, French) public research university in Canada. International students from 155 countries compose 17% of the student population at this institution.[15,16] We received ethics approval through the institution's Research Ethics Board (#H-01-22-7584).

We addressed the research questions *via* qualitative focus groups and interviews. Qualitative research allows participants to provide rich, detailed accounts of their experiences through methods like interviews and focus groups.[17] Qualitative researchers also often choose participants purposefully to address specific research questions. Though qualitative studies are limited in their *statistical-probabilistic generalizability*, this type of generalizability is inherently not the goal of qualitative work.[17,18] Rather, qualitative researchers often seek to establish *naturalistic generalizability*, defined as "generalizability that is reached on the basis of recognition of similarities and differences to the results with which the reader is familiar",[19] as well as *transferability*, which is achieved when "readers feel as though the story of the research overlaps with their own situation and they intuitively transfer the research to their own action."[20] This distinction is also why we do not provide frequencies for themes uncovered in the study.

The study sample (N = 18) was composed of 16 graduate students and 2 post-doctoral researchers (PDRs), referred to as *trainees* for simplicity (also called high-qualified personnel, HQP, in other work). We did not distinguish between the experiences of graduate students and post-doctoral researchers in this study. Participants were involved in research from a range of chemistry subdisciplines (*e.g.*, flow chemistry, biochemistry, organometallic, computational, organic), academic stages (3 MSc, 13 PhD, 2 PDR), and language backgrounds (Figure 1 in main article). All participants identified as having learned English-as-an-additional language. Of the participants, 16 identified as international trainees; 2 had also completed their undergraduate education in Québec, Canada in which French is the dominant language.

We recruited participants *via* their department's email listserv. The recruitment text invited trainees to participate in the study and directed them to a pre-session questionnaire. The pre-session questionnaire prompted participants to provide their contact information, availability, if they wanted a translator at their session, and to identify challenges Eng+ trainees might face when learn, communicating, and doing chemistry in English (Figure S1). To ensure participants felt as comfortable as possible during the sessions, we offered participants their choice between a focus group (up to 5 participants) or a 1-on-1 interview (or they could indicate that they had no preference). Participants were offered $20 gift cards as compensation.

We developed a semi-structured interview and focus group protocol (provided later in this SI) based on the research questions, qualitative methods literature,[21–27] and previous literature on Eng+ scientists.[28–31] If participants indicated specific topics they wished to discuss in the pre-session questionnaire, we used these topics as a starting point for the interview or focus group. We used the same protocol for both the focus groups and interviews, though participants in the focus groups were explicitly encouraged to respond to each other, ask each other questions, and discuss questions and topics as a group.[22,24,32] A total of three focus groups (each with three participants, average of 90 minutes) and nine interviews (average of 60 minutes) were conducted. All sessions were conducted online due to the COVID-19 pandemic and audio- and video-recorded. J.M.D. facilitated and took notes during each session. Immediately after the sessions, participants were invited to complete a post-session demographic questionnaire to collect data on different aspects of their

sociodemographic identities, including gender, race/ethnicity, and language background.[33,34] Both the pre- and post-session questionnaires are provided later in this SI. All participants were invited to provide preferred pseudonyms to be used when communicating this research (including this article).

Data analysis

We bring rigour to this work by grounding it in established methods for qualitative analysis.[35,36] First, audio data from the sessions were transcribed verbatim to generate transcripts. This was followed by line-by-line coding of the transcript data. The goal of line-by-line coding is to break down the data into discrete parts (ideas, perspectives, *etc.*) to facilitate examination and comparison of the data.[35,36] We focused on identifying distinct ideas within the transcripts, ranging in length from sentence fragments to statements up to four sentences long. For example, the quote "I think profs need to understand that just because you can speak it doesn't mean you can write it properly. I think a writing course would be great. A mandatory writing course." was coded as *Believes a mandatory writing course would help Eng+ writing, even those who seem to speak English well.* After identifying 1811 initial codes from line-by-line coding, pattern coding was employed to identify smaller sets of categories.[35,36] For example, two initial codes *Worried about looking dumb if they struggle to connect words in English* and *Doesn't speak up at group meetings for fear of being misunderstood and criticized* were grouped under the broader category *Worried about being viewed as less knowledgeable by colleagues.* After identifying broader categories, we organized the data in terms of the principal research questions, thereby identifying overarching themes that address the research questions. Qualitative coding is an iterative process—we consulted relevant literature throughout the process and critically discussed our analysis and themes with each other and other members of our research group.[37] Member-checking was used to maximize the validity of the themes we uncovered, ensuring that they resonated with participants' lived experiences as closely as possible. Sharing our preliminary codebook with participants, we asked participants (1) whether the themes we uncovered resonated with their experiences (or not), and (2) if there were any experiences they felt were missing from the themes and their descriptions.[38] Participants who responded to our member-checking invitation felt that the themes adequately represented their experiences, with some participants suggesting additional examples from their personal experiences to be added/highlight in the descriptors for each theme.

Following member-checking, reliability of the analysis was evaluated *via* inter-rater reliability.[39–41] Author J.M.D. and a second coder independently reviewed two transcripts using our codebook to identify the presence/absence of themes within each transcript. Following this, the two raters met to discuss any disagreements between their respective analyses; there were minor disagreements between the two coders, which resulted in some minor updates to theme definitions for clarity or combining redundant themes. For example, RQ3 theme 3 initially included a subtheme called *Aligning expectations with students' goals*. After discussion, both J.M.D. and the second coder found that this subtheme could be incorporated into *Providing trainees with resources and supports*. Full transcripts are not provided in the SI to preserve participants' anonymity.

REFERENCES

(1) Lee, O.; Stephens, A. English Learners in STEM Subjects: Contemporary Views on STEM Subjects and Language With English Learners. *Educ. Res.* **2020**, *49* (6), 426–432. https://doi.org/10.3102/0013189X20923708.
(2) Grapin, S. E.; Llosa, L.; Haas, A.; Lee, O. Rethinking Instructional Strategies with English Learners in the Content Areas. *TESOL J.* **2020**, 1–12. https://doi.org/10.1002/tesj.557.
(3) Lee, O. Asset-Oriented Framing of Science and Language Learning with Multilingual Learners. **2021**, No. April, 1–7. https://doi.org/10.1002/tea.21694.
(4) Lee, O.; Quinn, H.; Valdés, G. Science and Language for English Language Learners in Relation to Next Generation Science Standards and with Implications for Common Core State Standards for English Language Arts and Mathematics. *Educ. Res.* **2013**, *42* (4), 223–233. https://doi.org/10.3102/0013189X13480524.
(5) The National Academy of Sciences; National Academies of Science Engineering and Medicine; The National Academy of Sciences. *English Learners in STEM Subjects: Transforming Classrooms, Schools, and Lives*; Francis, D., Stephens, A., Eds.; The National Academies Press: Washington, D.C., 2018. https://doi.org/10.17226/25182.
(6) Grapin, S. E. Toward Asset-Oriented and Definitionally Clear Terminology: A Comment on González-Howard and Suárez (2021). *J. Res. Sci. Teach.* **2021**, *58* (5), 753–755. https://doi.org/10.1002/tea.21686.
(7) Yore, L. D.; Hand, B.; Goldman, S. R.; Hildebrand, G. M.; Osborne, J. F.; Treagust, D. F.; Wallace, C. S. New Directions in Language and Science Education Research. *Read. Res. Q.* **2004**, *39* (3), 347–352. https://doi.org/10.1598/RRQ.39.3.8.
(8) Afitska, O.; Heaton, T. J. Mitigating the Effect of Language in the Assessment of Science: A Study of English-language Learners in Primary Classrooms in the United Kingdom. *Sci. Educ.* **2019**, 1396–1422. https://doi.org/10.1002/sce.21545.
(9) Solano-Flores, G.; Trumbull, E. Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners. *Educ. Res.* **2003**, *32* (2), 3–13. https://doi.org/10.3102/0013189X032002003.
(10) Abedi, J. Assessment and Accommodations of English Language: Issues, Concerns, and Recommendations. *J. Sch. Improv.* **2002**, *3* (1), 83–89.
(11) Wolf, M. K.; Kao, J. C.; Griffin, N.; Herman, J.; Bachman, P. L.; Chang, S. M.; Farnsworth, T. *Issues in Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses*; 2008.
(12) LaCosse, J.; Canning, E. A.; Bowman, N. A.; Murphy, M. C.; Logel, C. A Social-Belonging Intervention Improves STEM Outcomes for Students Who Speak English as a Second Language. *Sci. Adv.* **2020**, *6* (40), 1–11. https://doi.org/10.1126/sciadv.abb6543.
(13) Abedi, J.; Zhang, Y.; Rowe, S. E.; Lee, H. Examining Effectiveness and Validity of Accommodations for English Language Learners in Mathematics: An Evidence-Based Computer Accommodation Decision System. *Educ. Meas. Issues Pract.* **2020**, *39* (4), 41–52. https://doi.org/10.1111/emip.12328.
(14) Buxton, C.; Allexsaht-Snider, M.; Aghasaleh, R.; Kayumova, S.; Kim, S.-H.; Choi, Y.-J.; Cohen, A. Potential Benefits of Bilingual Constructed Response Science Assessments for Understanding Bilingual Learners' Emergent Use of Language of Scientific Investigation Practices. *Double Helix* **2014**, *2* (September 2016), 1–21.
(15) University of Ottawa. Quick Facts 2020 https://www.uottawa.ca/institutional-research-planning/resources/facts-figures/quick-facts (accessed May 10, 2022).
(16) University of Ottawa. *Programs and Admissions*; 2021.
(17) Braun, V.; Clarke, V. *Successful Qualitative Research: A Practical Guide for Beginners*; Sage Publications Inc.: London, 2013.
(18) Smith, B. Generalizability in Qualitative Research: Misunderstandings, Opportunities and Recommendations for the Sport and Exercise Sciences. *Qual. Res. Sport. Exerc. Heal.* **2018**, *10* (1), 137–149. https://doi.org/10.1080/2159676X.2017.1393221.
(19) Stake, R. E. The Case Study Method in Social Inquiry. *Educ. Res.* **1978**, *7* (2), 5–8. https://doi.org/10.3102/0013189X007002005.
(20) Tracy, S. J. Qualitative Quality: Eight a"big-Tent" Criteria for Excellent Qualitative Research. *Qual. Inq.* **2010**, *16* (10), 837–851. https://doi.org/10.1177/1077800410383121.
(21) Morgan, D. L.; Hoffman, K. Focus Groups. In *The SAGE Handbook of Qualitative Data Collection Focus Groups*; 2018; pp 250–263.
(22) Tynan, A. C.; Drayton, J. L. Conducting Focus Groups—a Guide for First-Time Users. *Mark. Intell. Plan.* **1988**, *6* (1), 5–9. https://doi.org/10.1108/eb045757.
(23) Morgan, D. Focus Groups. *Annu. Rev. Sociol.* **1996**, *22*, 129–152.
(24) Onwuegbuzie, A. J.; Dickinson, W. B.; Leech, N. L.; Zoran, A. G. A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. *Int. J. Qual. Methods* **2009**, *8* (3), 1–21. https://doi.org/10.1177/160940690900800301.
(25) Cresswell, J. W. Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research; Pearson: Boston, MA, 2012.
(26) Patton, M. Q. Distinguishing Interview Approaches and Types of Interviews. In *Qualitative Research & Evaluation Methods*; 2015; pp 641–648.
(27) Welch, J. K.; Patton, M. Q. Qualitative Evaluation and Research Methods. *Mod. Lang. J.* **1992**. https://doi.org/10.2307/330063.
(28) Ramírez-Castañeda, V. Disadvantages in Preparing and Publishing Scientific Papers Caused by the Dominance of the English Language in Science: The Case of Colombian Researchers in Biological Sciences. *PLoS One* **2020**, *15* (9), 1–15. https://doi.org/10.1371/journal.pone.0238372.
(29) Elnathan, R. English is the language of science — but precision is tough as a non-native speaker https://www.nature.com/articles/d41586-021-00899-y.
(30) Ortega, R. P. Science's English dominance hinders diversity—but the community can work toward change https://www.sciencemag.org/careers/2020/10/science-s-english-dominance-hinders-diversity-community-can-work-toward-change (accessed Oct 30, 2020).
(31) Lee, O.; Fradd, S. H. Science for All, Including Students from Non-English-Language Backgrounds. *Educ. Res.* **1998**, *27* (4), 12–21. https://doi.org/10.3102/0013189X027004012.
(32) Twinn, S. An Analysis of the Effectiveness of Focus Groups as a Method of Qualitative Data Collection with Chinese Populations in Nursing Research. *J. Adv. Nurs.* **1998**, *28* (3), 654–661. https://doi.org/10.1046/j.1365-2648.1998.00708.x.
(33) Marian, V.; Blumenfeld, H. K.; Kaushanskaya, M. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *J. Speech, Lang. Hear. Res.* **2007**, *50* (4), 940–967. https://doi.org/10.1044/1092-4388(2007/067).
(34) Kaushanskaya, M.; Blumenfeld, H. K.; Marian, V. The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten Years Later. *Bilingualism* **2019**, 8–13. https://doi.org/10.1017/S1366728919000038.
(35) Saldaña, J. *The Coding Manual for Qualitative Researchers*, 3rd ed.; Sage Publications Inc.: Thousand Oaks, CA, 2016.
(36) Bazeley, P.; Jackson, K. *Qualitative Data Analysis with NVivo*; Sage Publications Inc.: Thousand Oaks, CA, 2013; Vol. 40. https://doi.org/10.1080/02607476.2013.866724.
(37) Kember, D.; Ha, T. S.; Lam, B. H.; Lee, A.; Ng, S.; Yan, L.; Yum, J. C. K. The Diverse Role of the Critical Friend in Supporting Educational Action Research Projects. *Educ. Action Res.* **1997**, *5* (3), 463–481. https://doi.org/10.1080/09650799700200036.
(38) Creswell, J.W., & Miller, D. L. Determining Validity in Qualitative Inquiry. *Theory Pract.* **2000**, *39* (3), 124–130. https://doi.org/10.1207/s15430421tip3903.

(39) Watts, F. M.; Finkenstaedt-Quinn, S. A. The Current State of Methods for Establishing Reliability in Qualitative Chemistry Education Research Articles. *Chem. Educ. Res. Pract.* **2021**, *22* (3), 565–578. https://doi.org/10.1039/d1rp00007a.
(40) Kenny, D. A. PERSON: A General Model of Interpersonal Perception. *Personal. Soc. Psychol. Rev.* **2004**, *8* (3), 265–280. https://doi.org/10.1207/s15327957pspr0803_3.
(41) Garrison, D. R.; Cleveland-Innes, M.; Koole, M.; Kappelman, J. Revisiting Methodological Issues in Transcript Analysis: Negotiated Coding and Reliability. *Internet High. Educ.* **2006**, *9* (1), 1–8. https://doi.org/10.1016/j.iheduc.2005.11.001.

DATA FROM PARTICIPANTS' PRE- AND POST-SESSION QUESTIONNAIRE RESPONSES

Table S1: Participants' pseudonyms and first language(s) ($N = 18$).

Participant	First language(s)
Xena	Persian (Farsi)
Paul	Arabic
Marjan	Persian (Farsi), Azerbaijani
Nima	Persian (Farsi)
Abbad	Arabic
Alicia	Russian, Tatar
Matar	French
Abita	French, Spanish
Stevie	German
Paula	French
Daniella	Portuguese
Sunny	Thai/Laotian
Noah	Persian (Farsi)
Kira	Arabic
Cecilia	Cantonese, Mandarin
Bill	French
Dave	Spanish
Jennifer	Hindi, Punjabi

Figure S1: Participants' responses to "In what situations do language barriers arise the most for you? Check all that apply."

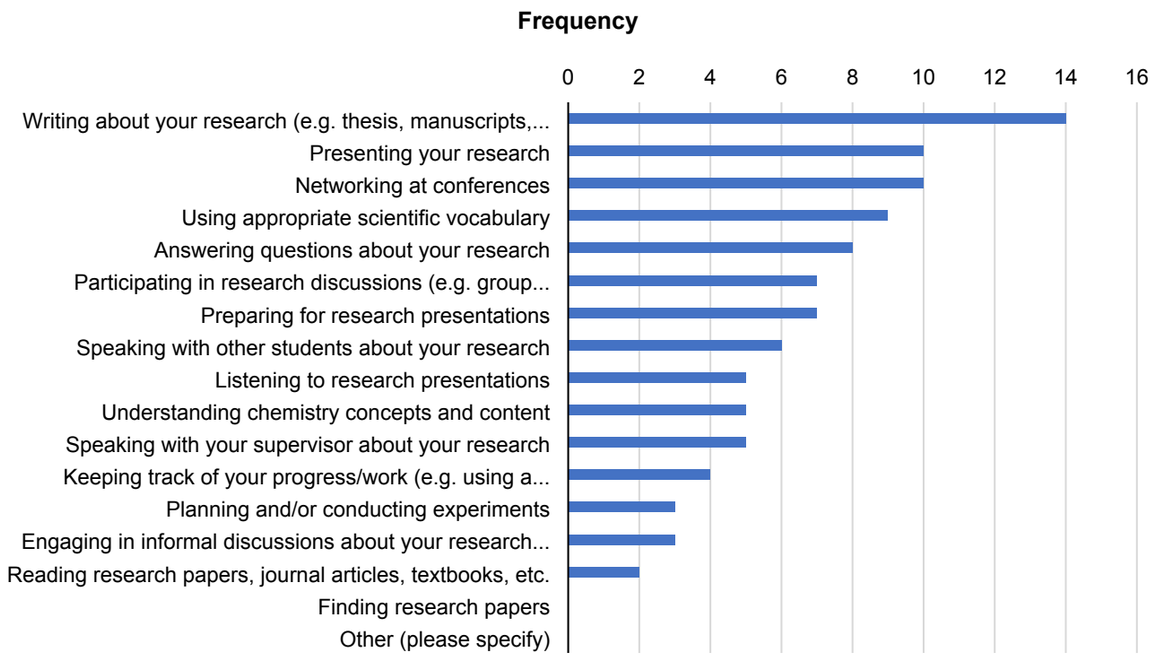


Figure S2: Participants' ethno-racial backgrounds ($N = 18$).

Note: We combined data for East Asian (e.g. Chinese, Japanese, Korean, Mongolian, Polynesian), South Asian (e.g. Indian, Pakistani, Sri Lankan, Bangladeshi), or Southeast Asian (e.g. Filipino, Burmese, Cambodian, Malaysian, Laotian, Thai, Vietnamese) to protect participant(s) anonymity. Participant who selected "You don't have an option…" self-identified as "Brown".

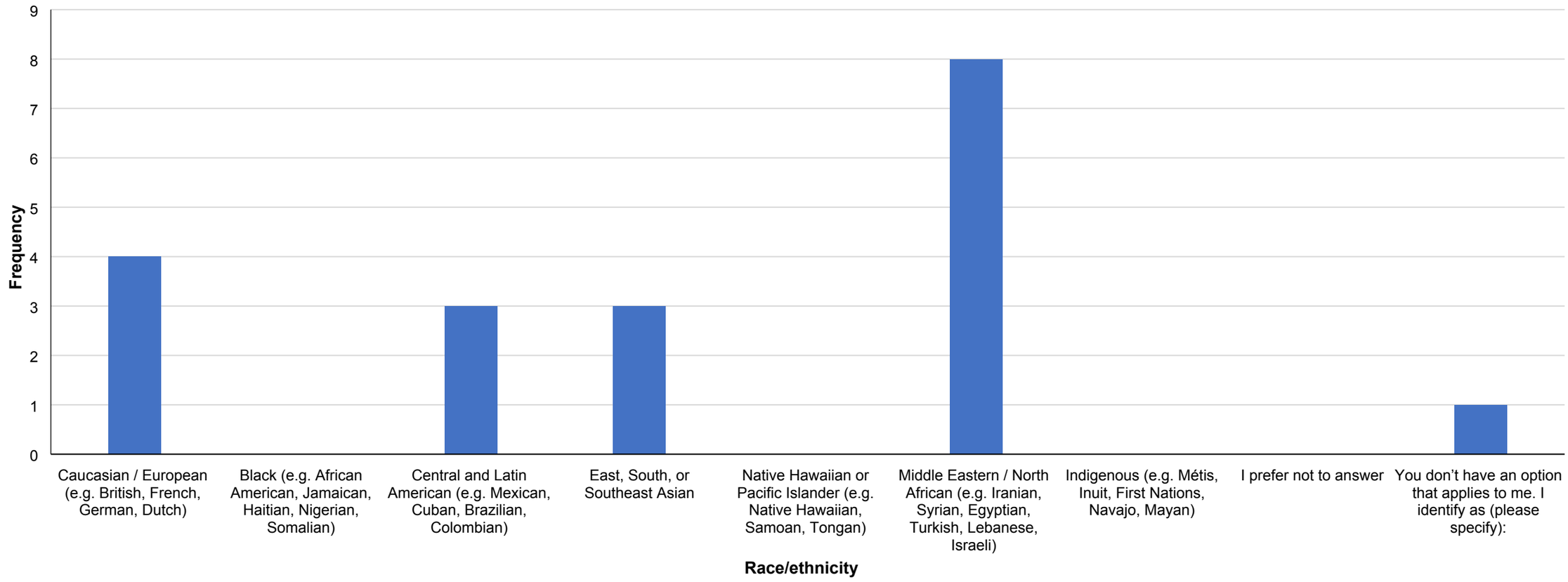


Figure S3: Participants' self-reported proficiencies understanding, writing, speaking, and reading in English ($N = 18$).

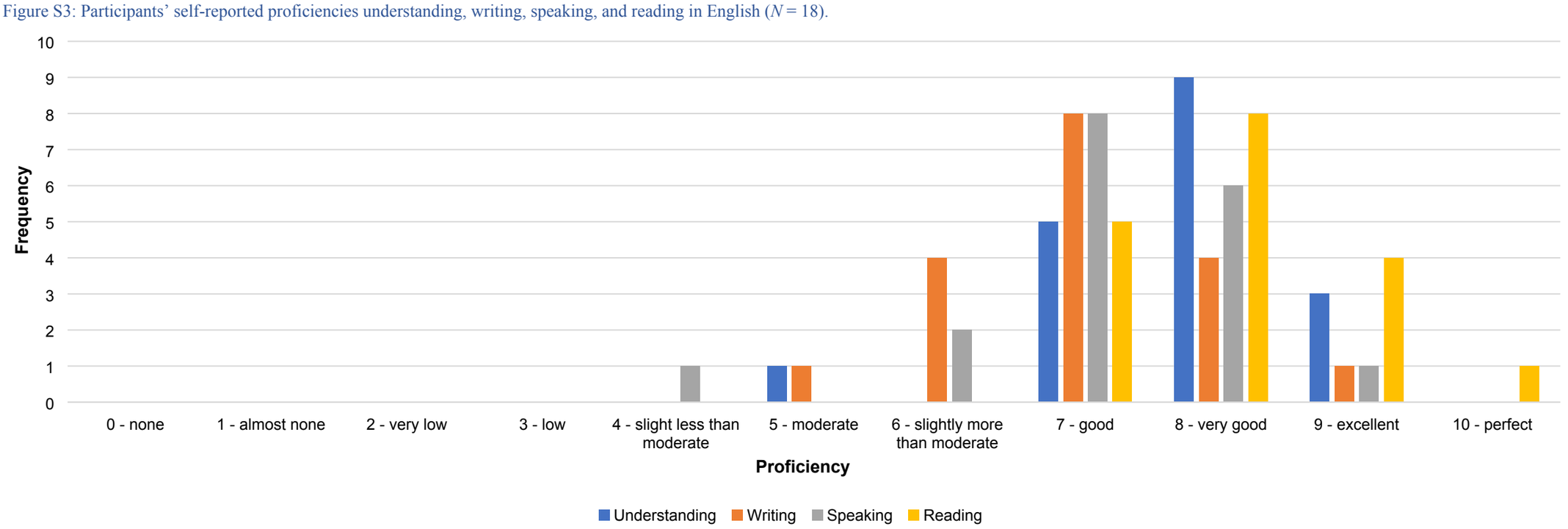


Figure S4: Participants' career stages, including years ($N = 18$).

Master's student = MSc; Doctoral student = PhD; Post-doctoral researcher = PDR.

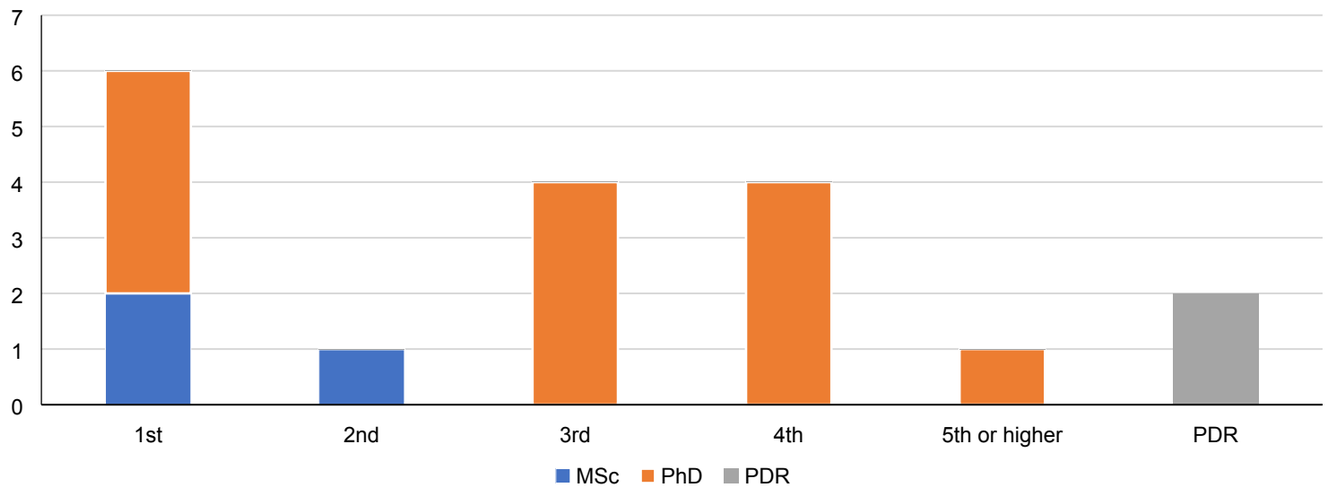
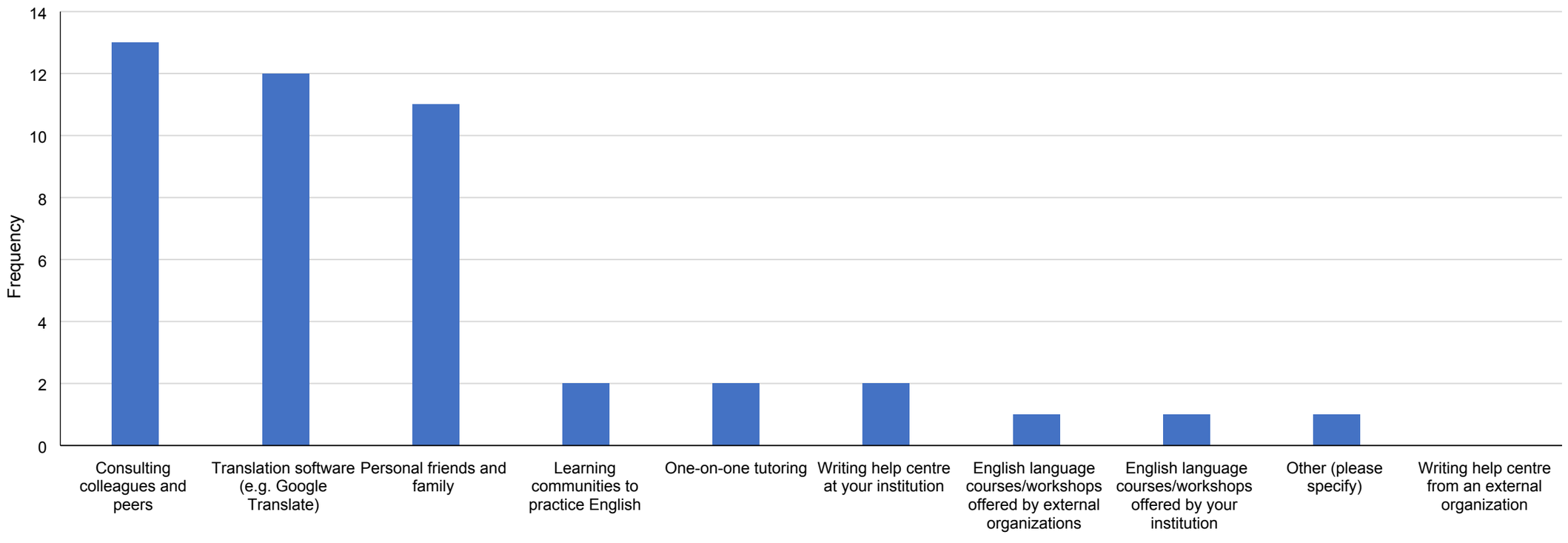


Figure S5: Participants' responses to: "What types of English language supports/services have you used or currently use? Check all that apply."



FLYNN RESEARCH GROUP

Eng+ chemistry grad students and PDFs - focus groups/interviews - pre

Consent form and sign-up

The purpose of this study is to determine how language shapes the research experiences of graduate students and post-doctoral fellows in chemistry. You will complete a brief questionnaire asking for your name, email, and experiences as a graduate student or postdoctoral fellow in chemistry prior to a focus group or interview. At the focus group/interview session, you participate in a recorded discussion of your experience as an Eng+ chemist. The researchers will then transcribe this data to identify key themes from your responses. The researchers may also invite you via email to review a summary of your contributions to the focus groups / interviews for feedback and/or clarity. After the focus group/interview, you will complete a short questionnaire asking to self-report your English language history and proficiency, as well as indicate whether the researchers can use your audio recordings in future research presentations and workshops. Selection will be on a first-come, first-served basis. After the intended number of participants is reached (N = 15), any additional participants expressing interest will be contacted via email to let them know that they have not been selected to participate and to ask them if they would like to be on a waiting list in case spots open.

You will be contributing to improving the quality of science education for yourself and your peers. You might also gain awareness of your own experiences as a graduate student/PDF, and how to better advocate for your experiences in the future. There are no risks associated with this study.

Participants' names, emails, and demographic data will be kept by the PI and PhD student. We will NOT share any of this information with anyone else at any time. We will ensure anonymity by referring to you using a pseudonym in all communications, including any quotes. Participating in this study will have no impact on your academic or career standing. All participants who participate in a focus group agree to keep all discussions during the focus group confidential. Survey data will be collected via SurveyMonkey. Survey questions and responses from respondents will be stored in SurveyMonkey's cloud storage based in Canada. All digital data (audio/video recording, survey responses) will be stored on the PI, CR, and RA's shared, private Google Drive folder. Only the PI can grant access to these folders beyond the research team. All members of the research team will have signed a confidentiality agreement. The data will be stored indefinitely.

You will be offered a $25 CAD Amazon or Tim Hortons gift (your choice) for your time. You are under no obligation to participate and acknowledge that participating is voluntary. Choosing to participate (or not) will have no effect on your academic standing or relationship with the professor conducting the project. If you choose to withdraw, you provide the researchers with consent to use data you contributed in any focus group discussions to preserve the integrity of the overall discussion data.

If you choose to withdraw, you are still entitled to receive the full compensation amount for participation in this study.

Please save/print a copy of this page for your records. If you have any questions about the study, you may contact Jacky Deng ([redacted]) or Alison Flynn [redacted] If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5 Tel.: (613) 562-5387 Email: ethics@uottawa.ca

By completing this form, you agree to participate in the above research study conducted by Jacky Deng of the Department of Chemistry and Biomolecular Sciences in the Faculty of Science at the University of Ottawa, which is under the supervision of Alison Flynn.

* 1. **What is your full name?**

* 2. **What is your uOttawa email?**

* 3. **In what situations do language barriers arise the most for you?** Check all that apply.

- [ ] Listening to research presentations
- [ ] Finding research papers
- [ ] Reading research papers, journal articles, textbooks, etc.
- [ ] Understanding chemistry concepts and content
- [ ] Using appropriate scientific vocabulary
- [ ] Planning and/or conducting experiments
- [ ] Keeping track of your progress/work (e.g. using a research notebook)
- [ ] Participating in research discussions (e.g. group meetings, conferences)
- [ ] Preparing for research presentations
- [ ] Presenting your research
- [ ] Answering questions about your research
- [ ] Writing about your research (e.g. thesis, manuscripts, etc.)
- [ ] Speaking with other students about your research
- [ ] Speaking with your supervisor about your research
- [ ] Networking at conferences
- [ ] Engaging in informal discussions about your research (e.g. elevator pitch)
- [ ] Other (please specify)

4. Optional: **Would you like to elaborate on any of your choices in the previous question?**

* 5. Would you prefer to participate in a **1-on-1 interview** or a **focus group (maximum 7 participants)?**

- ◯ Interview
- ◯ Focus group
- ◯ No preference (whatever is available)
- ◯ Both

* 6. **Please check all days/times you would be available for a focus group/interview.** Sessions will take a maximum of 90 minutes and you can leave the session at any time.

- ☐ Mon, Mar 14, 9:30-11:00
- ☐ Mon, Mar 14, 13:00-14:30
- ☐ Mon, Mar 14, 14:30-16:00
- ☐ Mon, Mar 14, 16:00-17:30
- ☐ Tues, Mar 15, 9:30-11:00
- ☐ Tues, Mar 15, 11:00-12:30
- ☐ Tues, Mar 15, 12:30-14:00
- ☐ Tues, Mar 15, 14:00-15:30
- ☐ Tues, Mar 15, 15:30-17:00
- ☐ Wed, Mar 16, 9:30-11:00
- ☐ Wed, Mar 16, 11:00-12:30
- ☐ Thurs, Mar 17, 9:30-11:00
- ☐ Thurs, Mar 17, 12:00-13:30
- ☐ Thurs, Mar 17, 13:30-15:00
- ☐ Thurs, Mar 17, 15:00-16:30
- ☐ Fri, Mar 18, 9:30-11:00
- ☐ Fri, Mar 18, 11:00-12:30
- ☐ Fri, Mar 18, 14:00-15:30
- ☐ Fri, Mar 18, 15:30-17:00
- ☐ Please let us know any additional times you are available

* 7. Would you like a **$25 CAD gift card** to **Amazon** or **Tim Hortons**?

- ◯ Amazon
- ◯ Tim Hortons

8. Optional: If you would like a **translator** present at the focus group/interview, **please indicate the language.** Note that this does not guarantee that a translator will be present.

FLYNN RESEARCH GROUP

Eng+ chemistry grad students and PDFs - focus groups/interviews - post

Wrap-up survey

Thank you for participating in the discussion. We invite you to complete the following demographic questions to wrap-up the study. Your responses will be confidential and will be used only to provide context to topics that were discussed during the session.

* 1. **Can we use your recorded audio in future settings (e.g., research presentations and workshops)?**

This audio would be used to hear your unique perspective, rather than having someone else read your words. In all cases, we would assign you a pseudonym and pitch-shift the audio to preserve anonymity.

- ◯ Yes, I consent
- ◯ No, I do not consent

* 2. What is your **full name**?

* 3. What **year of graduate school** are you currently enrolled in?

- ◯ 1st
- ◯ 2nd
- ◯ 3rd
- ◯ 4th
- ◯ 5th or higher
- ◯ Postdoctoral researcher
- ◯ Other (please specify)

* 4. What **type of degree** are you pursuing? (e.g. Master's in Chemistry)

* 5. What **keywords** would you use to describe your research?

* 6. **What do you consider your first language(s)**? A first language is defined as a language you have learned since birth and continue to use today.

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other (please specify)

* 7. Which language(s) did you do your **undergraduate chemistry** in?

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other (please specify)

* 8. What language(s) do you conduct your **graduate/post-doctoral research** in?

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other
- [ ] Other (please specify)

* 9. How often do you use **English at home**?

0 - never 100 - always

* 10. How often do you use **English at school/work**?

0 - never 100 - always

* 11. How many **years** have you have spent in a...

Country where English is predominantly spoken?

Home/family where English is predominantly spoken?

School/workplace where English is predominantly spoken?

* 12. **What is your proficiency in...**

	0 - none	1 - almost none	2 - very low	3 - low	4 - slightly less than moderate	5 - moderate	6 - slightly more than moderate	7 - good	8 - very good	9 - excellent	10 - perfect
Understanding English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Writing in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaking in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Reading in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

* 13. What types of **English language supports/services** have you used or currently use? Check all that apply.

- ☐ Consulting colleagues and peers
- ☐ English language courses/workshops offered by external organizations
- ☐ English language courses/workshops offered by your institution
- ☐ Learning communities to practice English
- ☐ One-on-one tutoring
- ☐ Personal friends and family
- ☐ Translation software (e.g. Google Translate)
- ☐ Writing help centre at your institution
- ☐ Writing help centre from an external organization
- ☐ Other (please specify)

14. Please answer this question with a **preferred pseudonym** you would like us to use. This will be used in place of your name in any research communications for this project.

* 15. Which of the following **racial/ethnic groups** do you identify with?

- [ ] Caucasian / European (e.g. British, French, German, Dutch)
- [ ] Black (e.g. African American, Jamaican, Haitian, Nigerian, Somalian)
- [ ] Central and Latin American (e.g. Mexican, Cuban, Brazilian, Colombian)
- [ ] East Asian (e.g. Chinese, Japanese, Korean, Mongolian, Polynesian)
- [ ] South Asian (e.g. Indian, Pakistani, Sri Lankan, Bangladeshi)
- [ ] Southeast Asian (e.g. Filipino, Burmese, Cambodian, Malaysian, Laotian, Thai, Vietnamese)
- [ ] Native Hawaiian or Pacific Islander (e.g. Native Hawaiian, Samoan, Tongan)
- [ ] Middle Eastern / North African (e.g. Iranian, Syrian, Egyptian, Turkish, Lebanese, Israeli)
- [ ] Indigenous (e.g. Métis, Inuit, First Nations, Navajo, Mayan)
- [ ] I prefer not to answer
- [ ] You don’t have an option that applies to me. I identify as (please specify):

* 16. “Disability” includes a range of physical and mental health issues, including invisible disabilities and it encompasses all types of disabilities. **Do you identify as a person with a disability?**

- ◯ Yes
- ◯ No
- ◯ I prefer not to answer
- ◯ You don’t have an option that applies to me. I identify as (please specify):

* 17. **What sex were you assigned at birth**, meaning on your original birth certificate?

- ◯ Male
- ◯ Female

* 18. Which best describes your **current gender identity**? Please select all boxes that apply to you.

- [ ] Man
- [ ] Woman
- [ ] Non-binary / gender-fluid
- [ ] Indigenous or other cultural identity (e.g. two-spirit)
- [ ] I prefer not to answer
- [ ] You don’t have an option that applies to me. I identify as (please specify):

* 19. Which best describes your **current sexual orientation**? Please select all boxes that apply to you.

- [ ] Heterosexual
- [ ] Gay
- [ ] Lesbian
- [ ] Bisexual
- [ ] Pansexual
- [ ] Asexual
- [ ] Queer
- [ ] Indigenous or other cultural identity (e.g. two-spirit)
- [ ] Questioning / not sure
- [ ] I prefer not to answer
- [ ] You don’t have an option that applies to me. I identify as (please specify):

Graduate students and post-doctoral fellows – focus groups and interviews protocol

Session Checklist
- 90 minutes minimum
- 5-7 participants
- Zoom recording enabled
- Pre-questionnaire completed
- Research notebook
- Translator (if requested)
- Post-session survey: https://www.surveymonkey.ca/r/7GLCX6D

RQ: How has language shaped Eng+ chemistry graduate students and post-doctoral fellows, and experiences in research including:
1. What are their language experiences prior to working in English chemistry research programs?
2. What barriers have they faced while learning, communicating, and doing chemistry?
3. What strategies and supports have they found to be helpful and/or needed?

Moderator tips:
- Create a safe space. Let participants bring up sensitive information on their own before addressing as a moderator.
- Facilitate conversation and make sure the RQs are being addressed, but most of the conversation should be driven by participants.
- Notice how questions / answers arise. What is the group's consensus? What is shared / different between participants? How do people talk about an issue?
- Do not lose sight of RQs.
- Be mindful of who is/is not participating and invite them into the conversation. (e.g., Once you [name] have commented, I would be interested in hearing from people who haven't commented yet [gives a bit of time for those ppl to reflect on if they want to add and what])

Confidentiality Discussion (5 minutes)

Thank you for being here today. We really appreciate you spending the time today to participate in our study. This discussion will take approximately 90 minutes. However, if you need to leave before then, that is completely fine.

I want to reiterate the section of the consent form regarding confidentiality and anonymity. As a participant in this study, your names, emails, and demographic data will be kept by myself and my supervisor, Dr. Alison Flynn. This information will remain strictly confidential, and we will NOT share any of this information with anyone else at any time. We'll ensure anonymity by referring to you with a pseudonym in all communications, including any quotes. Participating in this study will have no impact on your academic or career standing. By participating in this discussion today, you also agree keep all discussions during the focus group confidential, including the names of other participants.

This session will be audio and video-recorded. These recordings will also us to transcribe and analyze all contributions to the discussion.

Main questions (60 minutes)

1. **Icebreaker activity:** We're going to start with a quick ice-breaker activity to get us warmed-up. We'll go around the room and introduce ourselves with our name, research focus, and something about you. I can start…

2. This study is about how we communicate in our day-to-day work as chemistry graduate students. We'll start with a brainstorm activity on the following question: **What are some ways we, chemistry graduate students, communicate our research?**
   a. If silent:
      i. Start with examples from pre-session survey
      ii. Is there anything missing here?

3. **Thanks for bringing that up! How do you approach this specific task/situation? For example:**
   a. How do you prepare for this situation? Is this different or the same for other situations?
   b. How frequently do you ask for feedback? From whom?
   c. How do these situations make you feel?
   d. How do you feel after these situations finish or when you complete the task

4. Everyone here today identified to us as someone who learned English-as-an-additional language. **How do you feel being Eng+ has affected your experience(s) in the situations we've discussed, if at all?**
   a. One claim that's been stated in the literature is _____. How does this compare against your experience?
      i. Negative: extra time needed, research progress, viewed negatively, publication bias, anxiety
      ii. Positive: new skills, excited to learn a new language

5. What do you recommend for other Eng+ grad students?

Negative

1. How does this make you feel?
   a. How long have you felt this way?
2. **How does this impact your research?**
3. **How does this impact your learning?**
4. **What has made / makes these situations easier?**
   a. Who/what do you turn to when you need help with this issue?
   b. What about these resources make these situations easier?
5. **Do you think you've developed any helpful skills or strategies because of these experiences?**
   a. If yes, what are they?
6. How widespread do you think this issue is?
   a. If widespread, what do you think could help?
7. How do you think your experience compares to someone who is Eng1st?

Positive

8. How does this make you feel?
   a. How long have you felt this way?
9. Was this perspective something you always held?
   a. If not, how did you view the situation before and what changed?
10. How widespread do you think this perspective is?
   a. If not, have you tried sharing your perspective with other Eng+ students?
11. **Do you think you've developed any helpful skills or strategies because of these experiences?**
   a. If yes, what are they?
12. How do you think your experience compares to someone who is Eng1st?

13. If participant shares **neutral experience**:
   a. Can you provide some more details on this experience?
   b. How long have you felt this way?

14. Do you help others who are Eng+ in writing/presenting/etc.?
    a. If yes, how do you approach these situations?
    b. If no, why not?

15. How can others (professors, other grad students, administrators, etc) support Eng+ grad students in chemistry?

16. What type of resources for Eng+ graduate students do you think would be helpful at uOttawa, current or missing?

    a. Have you used any of the existing resources? Were these helpful/unhelpful? Why?

Wrap-up

We'll wrap-up the session here. The last thing I'll ask you to today is fill-out the following post-session survey (https://www.surveymonkey.ca/r/VLMP9MD), which includes a few questions on your background and your experience today. Please let me know once you've completed the survey.

I'll be sending each of you your $25 gift cards in the days following this session.

At the end of today's session, you'll also have a chance to let us know if we can use your audio recording in presentations.

If you have any further questions, you may contact either my research advisor (Dr. Flynn) or me via email, which I have put in the chat.

I can stick around for a few minutes if anyone has any questions. Thanks, everyone.

13.4 Appendix 4: Chapter 7 Supporting Information

FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

Language selection

Thank you for participating in our study. This study's purpose is to determine how the need to use English in chemistry shapes the experiences of researchers from diverse language backgrounds. Our research questions are:

1. **How does the need to use English in science impact the research experiences and productivity of chemistry researchers from diverse language backgrounds?**
2. **What resources and strategies do chemistry researchers use to navigate language barriers during their research?**
3. **To what extent are chemistry researchers from diverse language backgrounds using generative artificial intelligence, if at all?**

The questionnaire is 3 pages long and will take about 20 minutes to complete.

—

Nous vous remercions de votre participation à notre étude. L'objectif de cette étude est de déterminer comment la nécessité d'utiliser l'anglais en chimie façonne les expériences des chercheurs de diverses origines linguistiques. Nos questions de recherche sont les suivantes :

1. **Comment la nécessité d'utiliser l'anglais dans le domaine scientifique influe-t-elle sur les expériences de recherche et la productivité des chercheurs en chimie d'origines linguistiques diverses ?**
2. **Quelles ressources et stratégies les chercheurs en chimie utilisent-ils pour surmonter les barrières linguistiques au cours de leurs recherches**
3. **Dans quelle mesure les chercheurs en chimie de diverses origines linguistiques utilisent-ils l'intelligence artificielle générative, si tant est qu'ils l'utilisent ?**

Le questionnaire comporte 3 pages et prendra environ 20 minutes à remplir.

1. Select the language you wish to complete the questionnaire in. / Sélectionnez la langue dans laquelle vous souhaitez répondre à le questionnaire.

- ◯ English / anglais
- ◯ French / français

FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

1 - Research experiences

Thank you for participating in our study. This study's purpose to determine how the need to use English in chemistry shapes the experiences of researchers from diverse language backgrounds. The questionnaire is 3 pages long and will take about 20 minutes to complete.

Participation: You will complete an online background questionnaire about your experiences learning, communicating, and doing chemistry in English, along with questions on your language and demographic background (~20 minutes).
Risks: None.
Benefits: You will be contributing to improving the quality of science education for yourself and your peers. You might also gain awareness of your experiences as a scientist from a diverse language background and how to better advocate for yourself and others in the future.
Confidentiality and anonymity: Participants' demographic data will be kept by the PI and PhD student. The information you share will remain strictly confidential. We will NOT share any of this information with anyone else at any time. We will ensure anonymity by referring to me using a pseudonym in all communications, including any quotes. Participating in this study will have no impact on your academic or career standing.
Conservation of data: Data will be collected via SurveyMonkey. Questions and responses from respondents will be stored in SurveyMonkey's cloud storage based in Canada. All digital data will be stored on the PI, CR, and RA's shared, private OneDrive folder. Only the PI can grant access to these folders beyond the research team. All members of the research team will have signed a confidentiality agreement. The data will be stored indefinitely.
Compensation: To thank you for your contribution to the research project, you will be entered into a draw for five $100 CAD Amazon gift cards. You can also share the questionnaire with others who fit the study criteria to receive an additional entry into the draw per person who completes the questionnaire based on your referral. Your contact information that you provide is collected for the purposes of contacting you for compensation. Your name and the contact information you provide will be kept confidential and then destroyed once the prizes have been awarded. We reserve the right to cancel the draw or cancel the awarding of the prize if the integrity of the draw or the research or the confidentiality of participants is compromised. The draw is governed by the applicable laws of Canada.
Voluntary participation: You are under no obligation to participate and if you choose to participate, you can withdraw from the study at any time and/or refuse to answer any questions, without suffering any negative consequences. If you choose to withdraw, all data gathered until the time you withdraw will be removed from the project. If you choose to withdraw, you are still entitled to receive the full compensation amount for participation in this study.
If you have any questions about the study, you may contact Jacky Deng ([redacted]) or Alison Flynn ([redacted]).
If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5
Tel.: (613) 562-5387; Email: ethics@uottawa.ca
Acceptance: By completing this questionnaire, you agree to participate in the above

research study conducted by Jacky Deng of the Department of Chemistry and Biomolecular Sciences in the Faculty of Science at the University of Ottawa, whose research is under the supervision of Alison Flynn.

* 2. How often do you experience each of the following situations?

	1 - never	2	3	4 - sometimes	5	6	7 - all the time
Difficulty reading research materials in English.	○	○	○	○	○	○	○
Difficulty understanding research discussions in English (e.g., talks, presentations).	○	○	○	○	○	○	○
Difficulty writing about my research in English (e.g., papers, abstracts, proposals).	○	○	○	○	○	○	○
Difficulty speaking about my research in English (e.g., talks, presentations, meetings, seminars).	○	○	○	○	○	○	○
Difficulty interacting with other researchers in English.	○	○	○	○	○	○	○

Other (please specify)

* 3. How often do you experience each of the following situations?

	1 - never	2	3	4 - sometimes	5	6	7 - all the time
Reading papers multiple times to understand the content.	○	○	○	○	○	○	○
Needing additional feedback/time on manuscripts and presentations prepared in English.	○	○	○	○	○	○	○
Concern that my manuscripts are more likely to be rejected because of my English ability.	○	○	○	○	○	○	○
Feeling anxious or overwhelmed when expressing my thoughts in English.	○	○	○	○	○	○	○
When I speak in English, it is a simplified or filtered version of my actual thoughts.	○	○	○	○	○	○	○
Difficulty responding to questions about my research (e.g. after a presentation).	○	○	○	○	○	○	○
Feeling like my English ability hinders my day-to-day research productivity.	○	○	○	○	○	○	○

Other (please specify)

* 4. How much do you disagree/agree with the following statements?

	1 - completely disagree	2	3	4 - neutral	5	6	7 - completely agree
Researchers with strong English skills have a better chance at publishing in high-impact research journals.	○	○	○	○	○	○	○
Researchers with weaker English skills are often mistaken for having poor research skills.	○	○	○	○	○	○	○
Researchers with weaker English skills deserve more time to prepare manuscripts, presentations, *etc*.	○	○	○	○	○	○	○

Other (please specify)

5. Is there anything else you would like to share regarding the need to use English in science? Please try to be as detailed as possible.

* 6. How helpful have you found each of the following strategies/resources for language support?

	1 - Not helpful at all	2	3 - Neutral	4	5 - Very helpful	I have not used this resource	This resource is not available to me
Support from my supervisor/mentor	○	○	○	○	○	○	○
Support from my peers and colleagues	○	○	○	○	○	○	○
Support from my personal friends and family	○	○	○	○	○	○	○
Artificial intelligence tools (e.g. ChatGPT)	○	○	○	○	○	○	○
Translation softwares (e.g. Google Translate, DeepL)	○	○	○	○	○	○	○
Writing assistant softwares (e.g. Grammarly)	○	○	○	○	○	○	○
Learning tools or apps (e.g. Duolingo)	○	○	○	○	○	○	○
Science-specific communication course	○	○	○	○	○	○	○
Academic writing centre at my institution	○	○	○	○	○	○	○
Translation services (from journals, institution, etc.)	○	○	○	○	○	○	○
English language courses/workshops outside of my institution	○	○	○	○	○	○	○
English language courses/workshops at my institution	○	○	○	○	○	○	○

Other (please specify)

* 7. Generative artificial intelligence (GenAI) is any artificial intelligence capable of generating text, images, or other media using generative models (e.g. ChatGPT).

In what ways have you primarily used generative artificial intelligence (e.g. ChatGPT) for language support, if at all?

- [ ] Translation
- [ ] Proofreading
- [ ] Content generation (e.g., creating written material)
- [ ] Language learning
- [ ] None
- [ ] Other (please specify)

* 8. Please explain why you have used GenAI for the tasks you chose in the previous question (Q8). If you chose "None", please explain why you have not used GenAI. Please try to be as detailed as possible.

9. What benefits or advantages could you see in using GenAI for language support? Please try to be as detailed as possible.

10. What limitations or challenges could you see in using GenAI for language support? Please try to be as detailed as possible.

11. How do you think others in the scientific community could better support scientists who are learning English, if at all? Please try to be as detailed as possible.



The role of English in chemistry research (ENG/FR)

2 - Language background

The following questions will ask you about your language background.

* 12. Which of the following statements best describes **your language background**?

- ◯ English IS one of my first languages, and I feel CONFIDENT using it to conduct research.
- ◯ English IS one of my first languages, but I STRUGGLE using it to conduct research.
- ◯ English IS NOT one of my first languages, but I feel CONFIDENT using it to conduct research.
- ◯ English IS NOT one of my first languages, and I STRUGGLE using it to conduct research.
- ◯ Other (please specify)

* 13. **What do you consider your first language(s)**? A first language is defined as a language you have learned since infancy and continue to use today.

- ☐ English
- ☐ French
- ☐ Arabic
- ☐ Armenian
- ☐ Atikamekw
- ☐ Bengali
- ☐ Cantonese
- ☐ Cree
- ☐ Dene
- ☐ Dutch
- ☐ German
- ☐ Greek
- ☐ Gujarati
- ☐ Haitian (French Creole)
- ☐ Hebrew
- ☐ Hindi
- ☐ Hmong
- ☐ Inuktitut
- ☐ Italian
- ☐ Japanese
- ☐ Korean
- ☐ Mandarin
- ☐ Mi'kmaq
- ☐ Montagnais (Innu)
- ☐ Navajo
- ☐ Oji-Cree
- ☐ Ojibway
- ☐ Persian (Farsi)
- ☐ Polish
- ☐ Portuguese
- ☐ Punjabi
- ☐ Romanian
- ☐ Russian
- ☐ Serbo-Croatian
- ☐ Spanish
- ☐ Tagalog
- ☐ Tamil
- ☐ Thai/Laotian
- ☐ Ukrainian
- ☐ Urdu
- ☐ Vietnamese
- ☐ Other (please specify)

* 14. What language(s) do you primarily use **in your daily life, outside of research (e.g. at home)**?

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other (please specify)

* 15. In what language(s) did you **learn chemistry in before starting research**? (*e.g.*, when you learned chemistry during high school and undergraduate studies)

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other (please specify)

* 16. What language(s) have you primarily used when **conducting research**?

- [ ] English
- [ ] French
- [ ] Arabic
- [ ] Armenian
- [ ] Atikamekw
- [ ] Bengali
- [ ] Cantonese
- [ ] Cree
- [ ] Dene
- [ ] Dutch
- [ ] German
- [ ] Greek
- [ ] Gujarati
- [ ] Haitian (French Creole)
- [ ] Hebrew
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] Italian
- [ ] Japanese
- [ ] Korean
- [ ] Mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persian (Farsi)
- [ ] Polish
- [ ] Portuguese
- [ ] Punjabi
- [ ] Romanian
- [ ] Russian
- [ ] Serbo-Croatian
- [ ] Spanish
- [ ] Tagalog
- [ ] Tamil
- [ ] Thai/Laotian
- [ ] Ukrainian
- [ ] Urdu
- [ ] Vietnamese
- [ ] Other (please specify)

* 17. How often do you use English during your **research/work**?

0 - never 100 - always

* 18. How often do you use English **outside of your research/work**?

0 - never 100 - always

* 19. How many years have you spent in... (please enter a whole number)

Countries where English is predominantly used?

Homes/family where English is predominantly used?

Schools/workplaces where English is predominantly used?

* 20. Compared to someone who has spoken English their entire life and considers it their first language, what is your **proficiency** with...

	0 - none	1 - almost none	2 - very low	3 - low	4 - slightly less than moderate	5 - moderate	6 - slightly more than moderate	7 - good	8 - very good	9 - excellent	10 - perfect
Understanding English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Writing in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaking in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Reading in English	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

FRG
FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

3 - Demographic questionnaire

The following questions will ask you about your research and demographic background.

* 21. What is your primary affiliation?

- ◯ Academic
- ◯ Government
- ◯ Industry (includes start-ups, private sector, etc.)
- ◯ Non-profit institution
- ◯ Retired
- ◯ Self-employed
- ◯ Unemployed
- ◯ Other (please specify)

22. If you selected "Academic" in the previous section, which best describes your position at your institution?

- ◯ Undergraduate student
- ◯ Graduate student
- ◯ Research assistant
- ◯ Postdoctoral researcher
- ◯ Assistant professor
- ◯ Associate professor
- ◯ Instructor (or equivalent)
- ◯ Professor
- ◯ Staff personnel (e.g. research technician, facility manager, etc.)
- ◯ Other (please specify)

* 23. In what country do you work?

* 24. How many years have you worked in research? Please enter a whole number (e.g., 1, 2, 3, 4...). No words, decimals, or percentages.

* 25. What are your primary field(s) of research in chemistry? Check all that apply.

- ☐ Analytical chemistry
- ☐ Biochemistry
- ☐ Chemical biology
- ☐ Chemistry education
- ☐ Chemical engineering
- ☐ Computational chemistry
- ☐ Electrochemistry
- ☐ Environmental chemistry
- ☐ Geochemistry
- ☐ Green chemistry
- ☐ Inorganic chemistry
- ☐ Materials chemistry
- ☐ Medicinal chemistry
- ☐ Nuclear chemistry
- ☐ Organic chemistry
- ☐ Physical chemistry
- ☐ Polymer chemistry
- ☐ Quantum chemistry
- ☐ Supramolecular chemistry
- ☐ Surface chemistry
- ☐ Theoretical chemistry
- ☐ Other (please specify)

* 26. Which of the following racial/ethnic groups do you identify with? Check all that apply.

- [ ] Black (e.g. African American, Jamaican, Haitian, Nigerian, Somalian)
- [ ] Caucasian / European (e.g. British, French, German, Dutch)
- [ ] Central and Latin American (e.g. Mexican, Cuban, Brazilian, Colombian)
- [ ] East Asian (e.g. Chinese, Japanese, Korean, Mongolian, Polynesian)
- [ ] Indigenous (e.g. Métis, Inuit, First Nations, Navajo, Mayan)
- [ ] Middle Eastern / North African (e.g. Iranian, Syrian, Egyptian, Turkish, Lebanese, Israeli)
- [ ] Native Hawaiian or Pacific Islander (e.g. Native Hawaiian, Samoan, Tongan)
- [ ] South Asian (e.g. Indian, Pakistani, Sri Lankan, Bangladeshi)
- [ ] Southeast Asian (e.g. Filipino, Burmese, Cambodian, Malaysian, Laotian, Thai, Vietnamese)
- [ ] I prefer not to answer
- [ ] You don’t have an option that applies to me. I identify as (please specify):

* 27. "Disability” includes a range of physical and mental health issues, including invisible disabilities and it encompasses all types of disabilities. **Do you identify as a person with a disability that affects your ability to communicate with others (e.g. speaking, writing, listening, understanding)?**

- ◯ Yes
- ◯ No
- ◯ I prefer not to answer
- ◯ You don’t have an option that applies to me. I identify as (please specify):

28. Do you have any feedback or additional thoughts about this study?

29. How did you hear about this study?

- [ ] Social media (Twitter / X, Facebook, Instagram, etc.)
- [ ] Email from a professional society
- [ ] Email from a student association
- [ ] Departmental/faculty email
- [ ] Referral from a friend, colleague, etc. Their email is:

FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

1 - Expériences de recherche

Nous vous remercions de votre participation à notre étude. L'objectif de cette étude est de déterminer comment la nécessité d'utiliser l'anglais en chimie façonne les expériences des chercheurs de diverses origines linguistiques. Le questionnaire comporte 3 pages et prendra environ 20 minutes à remplir.

Participation : Vous répondrez à un questionnaire en ligne sur mes expériences en matière d'apprentissage, de communication et de chimie en anglais, ainsi qu'à des questions sur mes antécédents linguistiques et démographiques (~20 minutes).
Risques : Aucun.
Avantages : Vous contribuerez à améliorer la qualité de l'enseignement des sciences pour moi et mes pairs. Vous pourrez également prendre conscience de vos expériences en tant que scientifique issu d'un milieu linguistique diversifié et apprendre à mieux défendre vos intérêts et ceux des autres à l'avenir.
Confidentialité et anonymat : Les données démographiques des participants seront conservés par le chercheur principal et le doctorant. Les informations que vous partagerez resteront strictement confidentielles. Nous ne partagerons AUCUNE de ces informations avec qui que ce soit, à aucun moment. Nous garantirons l'anonymat en me désignant par un pseudonyme dans toutes les communications, y compris les citations. La participation à cette étude n'aura aucune incidence sur vos études ou votre carrière.
Conservation des données : Les données seront collectées via SurveyMonkey. Les questions de le questionnaire et les réponses des personnes interrogées seront stockées dans le système de stockage en nuage de SurveyMonkey, basé au Canada. Toutes les données numériques seront stockées dans le dossier OneDrive privé et partagé du chercheur principal, du chercheur principal et du chercheur principal. Seul le chercheur principal peut autoriser l'accès à ces dossiers à d'autres membres de l'équipe de recherche. Tous les membres de l'équipe de recherche auront signé un accord de confidentialité. Les données seront stockées indéfiniment.
Rémunération : Pour vous remercier de votre contribution au projet de recherche, vous participerez à un tirage au sort de cinq cartes-cadeaux Amazon d'une valeur de 100 dollars canadiens.
Participation volontaire : Vous n'êtes pas obligé de participer et si vous choisissez de participer, vous pouvez vous retirer de l'étude à tout moment et/ou refuser de répondre à toute question, sans subir de conséquences négatives. Si vous choisissez de vous retirer, toutes les données recueillies jusqu'au moment de votre retrait seront supprimées du projet. Si vous choisissez de vous retirer, vous avez toujours le droit de recevoir le montant total de la compensation pour votre participation à cette étude.
Si vous avez des questions sur l'étude, vous pouvez contacter Jacky Deng ([redacted]) ou Alison Flynn ([redacted]).
Si vous avez des questions concernant la conduite éthique de cette étude, vous

pouvez contacter le Protocol Officer for Ethics in Research, Université d'Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5
Téléphone : (613) 562-5387 ; Courriel : ethics@uottawa.ca
Acceptation : En répondant à ce questionnaire, vous acceptez de participer à l'étude de recherche susmentionnée menée par Jacky Deng du département de chimie et des sciences biomoléculaires de la faculté des sciences de l'Université d'Ottawa, dont les recherches sont supervisées par Alison Flynn.

* 30. Quelle est la fréquence de chacune des situations suivantes ?

	1 - jamais	2	3	4 - parfois	5	6	7 - tout le temps
Difficulté à lire des documents de recherche en anglais.	○	○	○	○	○	○	○
Difficulté à comprendre les discussions de recherche en anglais (par exemple, les conférences, les présentations).	○	○	○	○	○	○	○
Difficulté à écrire sur mes recherches en anglais (par exemple, articles, résumés, propositions).	○	○	○	○	○	○	○
Difficulté à parler de mes recherches en anglais (par exemple, exposés, présentations, réunions, séminaires).	○	○	○	○	○	○	○
Difficulté à interagir avec d'autres chercheurs en anglais.	○	○	○	○	○	○	○

Autre (à préciser)

* 31. Quelle est la fréquence de chacune des situations suivantes ?

	1 - jamais	2	3	4 - parfois	5	6	7 - tout le temps
Lire plusieurs fois les documents pour en comprendre le contenu.	○	○	○	○	○	○	○
Besoin de plus de temps et de commentaires sur les manuscrits et les présentations préparés en anglais.	○	○	○	○	○	○	○
Je crains que mes manuscrits soient plus susceptibles d'être rejetés en raison de mes compétences en anglais.	○	○	○	○	○	○	○
Je me sens anxieux ou dépassé lorsque j'exprime mes idées en anglais.	○	○	○	○	○	○	○
Lorsque je m'exprime en anglais, il s'agit d'une version simplifiée ou filtrée de mes pensées réelles.	○	○	○	○	○	○	○
Difficulté à répondre aux questions sur mes recherches (par exemple, après une présentation).	○	○	○	○	○	○	○
J'ai l'impression que mes compétences en anglais entravent la productivité de mes recherches au quotidien.	○	○	○	○	○	○	○

Autre (à préciser)

* 32. Dans quelle mesure êtes-vous en désaccord ou en accord avec les affirmations suivantes ?

	1 - pas du tout d'accord	2	3	4 - neutre	5	6	7 - tout à fait d'accord
Les chercheurs qui maîtrisent l'anglais ont plus de chances de publier dans des revues de recherche à fort impact.	○	○	○	○	○	○	○
Les chercheurs ayant de faibles compétences en anglais sont souvent considérés à tort comme ayant de faibles compétences en recherche.	○	○	○	○	○	○	○
Les chercheurs qui maîtrisent moins bien l'anglais méritent plus de temps pour préparer leurs manuscrits, leurs présentations, etc.	○	○	○	○	○	○	○

Autre (à préciser)

33. Aimeriez-vous partager autre chose concernant la nécessité d'utiliser l'anglais dans les sciences ? Essayez d'être aussi détaillé que possible.

* 34. Dans quelle mesure avez-vous trouvé utile chacune des stratégies/ressources suivantes pour le soutien linguistique, si tant est qu'elles l'aient été ?

	1 - pas du tout utile	2	3 - Neutral	4	5 - très utile	Je n'ai pas utilisé cette ressource	Cette ressource n'est pas disponible pour moi
Soutien de votre superviseur/mentor	○	○	○	○	○	○	○
Soutien de vos pairs et collègues	○	○	○	○	○	○	○
Soutien de vos amis et de votre famille	○	○	○	○	○	○	○
Outils d'intelligence artificielle (par exemple ChatGPT)	○	○	○	○	○	○	○
Logiciels de traduction (par exemple Google Translate, DeepL)	○	○	○	○	○	○	○
Logiciels d'aide à la rédaction (par exemple Grammarly)	○	○	○	○	○	○	○
Outils ou applications d'apprentissage (par exemple Duolingo)	○	○	○	○	○	○	○
Cours de communication spécifique aux sciences	○	○	○	○	○	○	○
Centre de rédaction universitaire de votre établissement	○	○	○	○	○	○	○
Services de traduction (de revues, d'institutions, etc.)	○	○	○	○	○	○	○
Cours/ateliers d'anglais en dehors de votre établissement	○	○	○	○	○	○	○
Cours/ateliers d'anglais dans votre établissement	○	○	○	○	○	○	○

Autre (à préciser)

* 35. L'intelligence artificielle générative (GenAI) est toute intelligence artificielle capable de générer du texte, des images ou d'autres médias à l'aide de modèles génératifs (p. ex. ChatGPT).

De quelle manière avez-vous principalement utilisé l'intelligence artificielle générative (par exemple ChatGPT) pour le soutien linguistique, le cas échéant ?

- [ ] Traduction
- [ ] Relecture
- [ ] Génération de contenu (par exemple, création de matériel écrit)
- [ ] Apprentissage des langues
- [ ] Aucun

Autre (à préciser)

* 36. Veuillez expliquer pourquoi vous avez utilisé l'intelligence artificielle générative pour les tâches que vous avez choisies à la question précédente (Q8). Si vous avez choisi "Aucun", veuillez expliquer pourquoi vous n'avez pas utilisé l'intelligence artificielle générative. Essayez d'être aussi détaillé que possible.

37. Quels avantages voyez-vous à l'utilisation de l'intelligence artificielle générative pour le soutien linguistique ? Essayez d'être aussi détaillé que possible.

38. Quelles sont les limites ou les défis que vous voyez dans l'utilisation de l'intelligence artificielle générative pour le soutien linguistique ? Essayez d'être aussi détaillé que possible.

39. Comment pensez-vous que les autres membres de la communauté scientifique pourraient mieux soutenir les scientifiques qui apprennent l'anglais, le cas échéant ? Essayez d'être aussi détaillé que possible.

FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

2 - Contexte linguistique

Les questions suivantes vous interrogeront sur vos antécédents linguistiques.

* 40. Parmi les affirmations suivantes, laquelle décrit le mieux vos connaissances linguistiques ?

- ◯ L'anglais est l'une de mes premières langues et je me sens à l'aise pour l'utiliser dans le cadre de mes recherches.
- ◯ L'anglais est l'une de mes premières langues, mais j'ai du mal à l'utiliser pour mener des recherches.
- ◯ L'anglais n'est pas l'une de mes langues maternelles, mais je me sens à l'aise pour l'utiliser dans le cadre de mes recherches.
- ◯ L'anglais n'est pas l'une de mes langues maternelles et j'ai du mal à l'utiliser pour mener des recherches.
- ◯ Autre (à préciser)

* 41. Que considérez-vous comme votre (vos) première(s) langue(s) ? Une première langue est définie comme une langue que vous avez apprise depuis votre enfance et que vous continuez à utiliser aujourd'hui.

- ☐ Anglais
- ☐ français
- ☐ Arabe
- ☐ Arménien
- ☐ Atikamekw
- ☐ bengali
- ☐ Cantonais
- ☐ cri
- ☐ Dene
- ☐ néerlandais
- ☐ Allemand
- ☐ Grec
- ☐ Gujarati
- ☐ Haïtien (créole français)
- ☐ hébraïque
- ☐ Hindi
- ☐ Hmong
- ☐ Inuktitut
- ☐ italien
- ☐ japonais
- ☐ coréen
- ☐ mandarin
- ☐ Mi'kmaq
- ☐ Montagnais (Innu)
- ☐ Navajo
- ☐ Oji-Cree
- ☐ Ojibway
- ☐ Persan (Farsi)
- ☐ Polonais
- ☐ portugais
- ☐ Pendjabi
- ☐ roumain
- ☐ russe
- ☐ Serbo-croate
- ☐ Espagnol
- ☐ tagalog
- ☐ tamoul
- ☐ Thaï/Laotien
- ☐ ukrainien
- ☐ Ourdou
- ☐ vietnamien
- ☐ Autre (à préciser)

* 42. Quelle(s) langue(s) utilisez-vous principalement dans votre vie quotidienne, en dehors de la recherche (par exemple à la maison) ?

- [ ] Anglais
- [ ] français
- [ ] Arabe
- [ ] Arménien
- [ ] Atikamekw
- [ ] bengali
- [ ] Cantonais
- [ ] cri
- [ ] Dene
- [ ] néerlandais
- [ ] Allemand
- [ ] Grec
- [ ] Gujarati
- [ ] Haïtien (créole français)
- [ ] hébraïque
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] italien
- [ ] japonais
- [ ] coréen
- [ ] mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persan (Farsi)
- [ ] Polonais
- [ ] portugais
- [ ] Pendjabi
- [ ] roumain
- [ ] russe
- [ ] Serbo-croate
- [ ] Espagnol
- [ ] tagalog
- [ ] tamoul
- [ ] Thaï/Laotien
- [ ] ukrainien
- [ ] Ourdou
- [ ] vietnamien
- [ ] Autre (à préciser)

* 43. Dans quelle(s) langue(s) avez-vous appris la chimie avant de commencer la recherche ? (par exemple, lorsque vous avez appris la chimie au cours de vos études secondaires et universitaires)

- [ ] Anglais
- [ ] français
- [ ] Arabe
- [ ] Arménien
- [ ] Atikamekw
- [ ] bengali
- [ ] Cantonais
- [ ] cri
- [ ] Dene
- [ ] néerlandais
- [ ] Allemand
- [ ] Grec
- [ ] Gujarati
- [ ] Haïtien (créole français)
- [ ] hébraïque
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] italien
- [ ] japonais
- [ ] coréen
- [ ] mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persan (Farsi)
- [ ] Polonais
- [ ] portugais
- [ ] Pendjabi
- [ ] roumain
- [ ] russe
- [ ] Serbo-croate
- [ ] Espagnol
- [ ] tagalog
- [ ] tamoul
- [ ] Thaï/Laotien
- [ ] ukrainien
- [ ] Ourdou
- [ ] vietnamien
- [ ] Autre (à préciser)

* 44. Quelle(s) langue(s) avez-vous principalement utilisée(s) dans le cadre de vos recherches ?

- [ ] Anglais
- [ ] français
- [ ] Arabe
- [ ] Arménien
- [ ] Atikamekw
- [ ] bengali
- [ ] Cantonais
- [ ] cri
- [ ] Dene
- [ ] néerlandais
- [ ] Allemand
- [ ] Grec
- [ ] Gujarati
- [ ] Haïtien (créole français)
- [ ] hébraïque
- [ ] Hindi
- [ ] Hmong
- [ ] Inuktitut
- [ ] italien
- [ ] japonais
- [ ] coréen
- [ ] mandarin
- [ ] Mi'kmaq
- [ ] Montagnais (Innu)
- [ ] Navajo
- [ ] Oji-Cree
- [ ] Ojibway
- [ ] Persan (Farsi)
- [ ] Polonais
- [ ] portugais
- [ ] Pendjabi
- [ ] roumain
- [ ] russe
- [ ] Serbo-croate
- [ ] Espagnol
- [ ] tagalog
- [ ] tamoul
- [ ] Thaï/Laotien
- [ ] ukrainien
- [ ] Ourdou
- [ ] vietnamien
- [ ] Autre (à préciser)

* 45. À quelle fréquence utilisez-vous l'anglais dans le cadre de vos recherches/travaux ?

0 - jamais 100 - tout le temps

* 46. À quelle fréquence utilisez-vous l'anglais en dehors de votre recherche/travail ?

0 - jamais 100 - tout le temps

* 47. Combien d'années avez-vous passé à... (veuillez indiquer un nombre entier)

Pays où l'anglais est majoritairement utilisé ?

Foyer/famille où l'anglais est principalement utilisé ?

Écoles/lieux de travail où l'anglais est principalement utilisé ?

* 48. Par rapport à quelqu'un qui a parlé anglais toute sa vie et qui considère que c'est sa première langue, quel est votre niveau de...

	0 - aucun	1 - presque aucune	2 - très faible	3 - faible	4 - un peu moins que modéré	5 - modéré	6 - un peu plus que modéré	7 - bon	8 - très bon	9 - excellent	10 - parfait
Comprendre l'anglais	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Écrire en anglais	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Parler en anglais	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Lire en anglais	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

FRG
FLYNN RESEARCH GROUP

The role of English in chemistry research (ENG/FR)

3 - Questionnaire démographique

Les questions suivantes vous interrogeront sur vos recherches et votre situation démographique.

* 49. Quelle est votre affiliation principale ?

- ◯ Universités
- ◯ Gouvernement
- ◯ Industrie (y compris les start-ups, le secteur privé, etc.)
- ◯ Institution à but non lucratif
- ◯ Retraité
- ◯ Indépendant
- ◯ Chômeur
- ◯ Autre (à préciser)

50. Si vous avez choisi "Universitaire" dans la section précédente, quelle est la description la plus appropriée de votre fonction au sein de votre établissement ?

- ◯ Étudiant de premier cycle
- ◯ Étudiant diplômé
- ◯ Assistant de recherche
- ◯ Chercheur postdoctoral
- ◯ Professeur assistant
- ◯ Professeur associé
- ◯ Instructeur (ou équivalent)
- ◯ Professeur
- ◯ Personnel (par exemple, technicien de recherche, gestionnaire des installations, etc.)
- ◯ Autre (à préciser)

* 51. Dans quel pays travaillez-vous ?

* 52. Depuis combien d'années travaillez-vous dans la recherche ? Veuillez saisir un nombre entier (par exemple, 1, 2, 3, 4...). Pas de mots, de décimales ou de pourcentages.

* 53. Quel(s) est (sont) votre(s) principal(aux) domaine(s) de recherche en chimie ? Cochez toutes les cases correspondantes.

- ☐ Chimie analytique
- ☐ Biochimie
- ☐ Biologie chimique
- ☐ Enseignement de la chimie
- ☐ Génie chimique
- ☐ Chimie computationnelle
- ☐ Électrochimie
- ☐ Chimie de l'environnement
- ☐ Géochimie
- ☐ Chimie verte
- ☐ Chimie inorganique
- ☐ Chimie des matériaux
- ☐ Chimie médicinale
- ☐ Chimie nucléaire
- ☐ Chimie organique
- ☐ Chimie physique
- ☐ Chimie des polymères
- ☐ Chimie quantique
- ☐ Chimie supramoléculaire
- ☐ Chimie des surfaces
- ☐ Chimie théorique
- ☐ Autre (à préciser)

* 54. Parmi les groupes raciaux/ethniques suivants, auquel vous identifiez-vous ? Cochez toutes les cases correspondantes.

- [ ] Noirs (par ex. afro-américains, jamaïcains, haïtiens, nigérians, somaliens)
- [ ] Caucasiens / Européens (par ex. Britanniques, Français, Allemands, Néerlandais)
- [ ] Amérique centrale et latine (par ex. Mexique, Cuba, Brésil, Colombie)
- [ ] Asie de l'Est (par ex. chinois, japonais, coréen, mongol, polynésien)
- [ ] Indigènes (par ex. Métis, Inuits, Premières nations, Navajo, Mayas)
- [ ] Moyen-Orient / Afrique du Nord (par ex. Iranien, Syrien, Égyptien, Turc, Libanais, Israélien)
- [ ] Natifs hawaïens ou insulaires du Pacifique (par ex. natifs hawaïens, samoans, tongiens)
- [ ] Asiatique du Sud (par ex. Indien, Pakistanais, Sri Lankais, Bangladais)
- [ ] Asie du Sud-Est (par ex. Philippins, Birmans, Cambodgiens, Malaisiens, Laotiens, Thaïlandais, Vietnamiens)
- [ ] Je préfère ne pas répondre
- [ ] Vous n'avez pas d'option qui s'applique à moi. Je m'identifie comme (veuillez préciser) :

* 55. Le terme "handicap" recouvre toute une série de problèmes de santé physique et mentale, y compris les handicaps invisibles, et englobe tous les types de handicaps. Vous identifiez-vous comme une personne souffrant d'un handicap qui affecte votre capacité à communiquer avec les autres (par exemple, parler, écrire, écouter, comprendre) ?

- ◯ Oui
- ◯ Non
- ◯ Je préfère ne pas répondre
- ◯ Vous n'avez pas d'option qui s'applique à moi. Je m'identifie comme (veuillez préciser) :

56. Avez-vous des commentaires ou des réflexions supplémentaires à propos de ce questionnaire ? Par exemple, la longueur, les questions à ajouter, à mettre à jour, à supprimer, etc.

57. Comment avez-vous entendu parler de cette étude ?

- [ ] Médias sociaux (Twitter / X, Facebook, Instagram, etc.)
- [ ] Courriel d'une société professionnelle
- [ ] Courriel d'une association d'étudiants
- [ ] Courriel du département/de la faculté
- [ ] Recommandation d'un ami, d'un collègue, etc. Leur adresse électronique est la suivante :

Seeking chemistry researchers from any career stage

for an online survey on the use of English in chemistry

Research questions



How does the need to use English in science impact the research experiences and productivity of chemistry researchers from diverse language backgrounds?



What resources and strategies do chemistry researchers use to navigate language barriers during their research?



To what extent are chemistry researchers from diverse language backgrounds using generative artificial intelligence, if at all?

Details and how to participate:

Chance to win one of five **$100 Amazon gift cards!**

FRG FLYNN RESEARCH GROUP uOttawa

PI: Alison Flynn Contact: jdeng085@uottawa.ca Approved by uOttawa Ethics H-01-22-7584. All data will be kept confidential.

Recherche des chercheurs en chimie à tous les stades de leur carrière

pour une enquête en ligne sur l'utilisation de l'anglais en chimie

Research questions



Comment la nécessité d'utiliser l'anglais dans le domaine scientifique influe-t-elle sur les expériences de recherche et la productivité des chercheurs en chimie issus de milieux linguistiques différents ?



Quelles ressources et stratégies les chercheurs en chimie utilisent-ils pour surmonter les barrières linguistiques au cours de leurs recherches ?



Dans quelle mesure les chercheurs en chimie de diverses origines linguistiques utilisent-ils l'intelligence artificielle générative, si tant est qu'ils l'utilisent ?

Détails et comment participer :

Une chance de gagner l'une des cinq cartes-cadeaux Amazon de 100 $.

FRG FLYNN RESEARCH GROUP uOttawa

PI: Alison Flynn Email: jdeng085@uottawa.ca Approuvé par le dossier d'éthique de l'Université d'Ottawa H-01-22-7584. Toutes les données seront gardées confidentielles.

13.5 Appendix 5: Chapter 8 Supporting Information

Supporting Information: Response process validity evidence in chemistry education research

Jacky M. Deng, Nicholas Streja, and Alison B. Flynn*

Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON, Canada, K1N 6N5

Review of chemistry education research articles discussing the collection of response process validity evidence (N = 43). __1

Alternative methods to investigate response process _____19

Literature review method_____20

Additional figures from literature review _____23

Conducting interviews Remotely _____24

REVIEW OF CHEMISTRY EDUCATION RESEARCH ARTICLES DISCUSSING THE COLLECTION OF RESPONSE PROCESS VALIDITY EVIDENCE (*N* = 43).

Citation	Instrument	New or pre-existing?	Sample size	Protocol described?	Analysis described?	Modifications described?
Mulford, D. R.; Robinson, W. R. An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. J. Chem. Educ. 2002, 79 (6), 739.	Chemistry Concepts Inventory (CCI)	Developed new	8	None reported.	None reported.	General. No modifications needed, as interviews showed that the students interpreted the questions on the CCI as intended.

Barbera, J.; Adams, W. K.; Wieman, C. E.; Perkins, K. K. Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. J. Chem. Educ. 2008, 85 (10), 1435–1439.	Colorado Learning Attitudes about Science Survey (CLASS)	Modified pre-existing	10	Specific. Retrospective interviews where students completed survey on paper, then interviewer read each statement and student asked to give response and explanation of their response.	None reported.	Specific. Authors described specific revisions to wording based on confusion about selected wording.
Reardon, R. F.; Traverse, M. A.; Feakes, D. A.; Gibbs, K. A.; Rohde, R. E. Discovering the Determinants of Chemistry Course Perceptions in Undergraduate Students. J. Chem. Educ. 2010, 87 (6), 643–646.	Chemistry Course Perceptions (CCP) instrument	Developed new	10	General. These interviews were transcribed and examined by the researchers for items that could be associated with their perceptions. From this, seven items designed to measure chemistry course perceptions were written. These items were taken back to the undergraduate chemistry students and their instructors to see whether they felt that the items would measure students' perceptions regarding chemistry.	General. None of the students or instructors objected to the content or wording of the items.	General. No modifications needed.
Stains, M.; Escriu-Sune, M.; De Santizo, M. L. M. A.; Sevian, H. Assessing Secondary and College Students' Implicit Assumptions about the Particulate Nature of Matter: Development and Validation of the Structure and Motion of Matter Survey. J. Chem. Educ. 2011, 88 (10), 1359–1365.	Structure and Motion of Matter (SAMM) Survey	Developed new	25	None reported.	General. Interviews analyzed for readability, alignment between response and explanation, and scoring scheme alignment with student thinking.	General. Questions made more explicit.
Oloruntegbe, K. O.; Ikpe, A. Ecocultural Factors in Students' Ability to Relate Science Concepts Learned at School and Experienced at Home: Implications for Chemistry Education. J. Chem. Educ. 2011, 88 (3), 266–271.	Instrument created for study	Developed new	40	None reported.	None reported.	General. No modifications needed, as "No appreciable differences between data collected and interviews"

Christian, B. N.; Yezierski, E. J. Development and Validation of an Instrument to Measure Student Knowledge Gains for Chemical and Physical Change for Grades 6–8. Chem. Educ. Res. Pr. 2012, 13 (3), 384–393.	Chemical and Physical Change Assessment (CPCA)	Developed new	7	Specific. "The semi-structured interviews lasted for approximately 15 min and were think out-loud style. Each interviewee was asked to describe his/her thought process for an item as well as interpret all answer choices focusing on language and comprehension for at least 10 items from the CPCA a-version. The items for each interview were chosen to provide each participant with variety of items initially scored as correct and incorrect while ensuring that each item on the assessment received multiple interpretations by different students. In addition, administration observations were taken into account during the revision process."	General. Interview data was coded and summarized with tables.	Specific. "The student interviews guided minor revisions to 5 items' stems, 13 answer choices, and a handful of formatting changes. The nature of the changes ranged from simple clarifications to reorganizing item stems to reduce misinterpretations."

Wren, D.; Barbera, J. Gathering Evidence for Validity during the Design, Development, and Qualitative Evaluation of Thermochemistry Concept Inventory Items J. Chem. Educ. 2013, 90 (12), 1590-1601. ; Wren, D.; Barbera, J. Psychometric Analysis of the Thermochemistry Concept Inventory. Chem. Educ. Res. Pr. 2014, 15 (3), 380–390.	Thermochemistry Concept Inventory (TCI)	Developed new	16	Specific. "Therefore, a retrospective semistructured think-aloud interview approach using probing questions was used. Retrospective interviews, occurring immediately after students' answering testing items, are most appropriate for evaluation of test items in a test-taking environment. Student participants were given a stack of ordered multiple-choice items on separate pieces of paper and instructed to answer items, in order, for approximately 15 min, with emphasis on thoughtfully answering items rather than completing all items. To ensure equal item coverage in interviews, item order was changed for each interview. During the interview, students were asked to read the item stem and explain their response process in answering each item. This included interpretation of item stems, item figures, and item responses. Probing questions by the researcher focused on identifying what conceptions students were using to choose their answer, or to eliminate responses."	Specific. Interview data transcribed verbatim and coded using NVivo. Full analysis presented in the Supporting Information.	Specific. Concerns with item wording, item option plausibility, item option ordering, and conceptions students were using to answer item options were the driving force for item revisions. Examples for students responses and items that required revisions presented in Supporting Information.
Brandriet, A.R.; Bretz, S.L. The Development of the Redox Concept Inventory as a Measure of Students' Symbolic and Particulate Redox Understandings and Confidence. J. Chem. Educ. 2014, 91(8), 1132-1144.	Redox Concept Inventory (ROXCI)	Developed new	15	None reported.	General. Informal analysis to "verify that students interpreted the questions as intended and to verify that students chose the right answers for the right reasons".	General. Removal of irrelevant items and wording changes

Schwartz, P.; Barbera, J. Evaluating the content and response process validity of data from the chemical concepts inventory. J. Chem. Educ. 2014, 91 (5), 630–640.	Chemistry Concepts Inventory (CCI)	Modified pre-existing	25	Specific. "All interviews were audio recorded and transcribed verbatim. During each interview, students read through and answered each multiple-choice item individually. Students circled their intended answer on the test paper provided and were asked by the interviewer why they chose their particular answer. Depending upon their initial response for each item, students were then asked follow-up questions to provide clarity. After all follow-up questions were complete for an individual item, the student went on to the next item."	Specific. "Response choices for each interview phase were ultimately combined and analyzed together. These response choices were organized by response patterns that arose from each interview. The first author analyzed each interview transcript in detail, coding the response process of each participant. The second author reviewed sample transcript sections for each item; coded sections were compared between the two authors. When discrepancies arose, the transcript and codes were discussed until agreement was reached. Each interview response was coded on the basis of whether a participant provided evidence of conceptual understanding, used the intended alternate conception, expressed confusion or misunderstanding about the item or response options, or used test-taking strategies. Coded responses across all interviews were grouped by common ideas communicated for an individual item. Where two or more participants used similar discrepant reasoning to answer an item either correctly or incorrectly, this discrepancy was labeled as a potential threat to the response process validity. Coded excerpts for an individual were reviewed for consistency within an item group; these comparisons provided further support for individual item functioning."	None reported, though authors acknowledge that modifying the instrument was beyond the scope of this work.
Villalta-Cerdas, A.; Sandi-Urena, S. Self-Explaining Effect in General Chemistry Instruction: Eliciting Overt Categorical Behaviours by Design. Chem. Educ. Res. Pr. 2014, 15 (4), 530–540. https://doi.org/10.1039/C3RP00172E.	Learning materials to support self-explaining	Developed new	2	Specific. Think-aloud, described in detail in appendix.	None reported.	General. No modifications needed.

Luxford, C. J.; Bretz, S. L. Development of the Bonding Representations Inventory to Identify Student Misconceptions about Covalent and Ionic Bonding Representations. J. Chem. Educ. 2014, 91 (3), 312–320.	Bond Representations Inventory (BRI)	Developed new	2	General. On the basis of responses from the experts' reviews, two student validation interviews, and item analysis of the pilot data, questions were removed or modified.	None reported.	Specific. For example, questions were eliminated that did not clearly focus upon one of the four misconceptions themes that emerged from the analysis of interviews. Questions were modified to ensure that each question included a representation of bonding, and for questions that were quite similar to one another, one of the "duplicates" was eliminated.

Ferrell, B.; Barbera, J. Analysis of Students' Self-Efficacy, Interest, and Effort Beliefs in General Chemistry. Chem. Educ. Res. Pract. 2015, 16 (2), 318–337.	Self-Efficacy, Interest, and Effort Beliefs in General Chemistry Scale	Developed new (modified items from established scales)	14 (two rounds, 9 and 5, respectively)	Specific. "All interviews took place in a private interview room to ensure both participant confidentiality and audio quality. Prior to starting the interview, each participant was informed about the purpose of the study, the interview procedure, and the protocols for confidentiality. Following that, the participants signed a consent form approved by the IRB. Since students interviewed in the fall 2013 had already completed the scales at the start of the semester, they were provided a copy of their original answer choices for each scale. Students who were interviewed in the spring of 2014 were asked to complete the scales prior to the start of the interview. All interviews were audio recorded. A verbal probing interview approach was used, whereby students were asked to read each item out-loud, explain the reasoning behind the answer choice they made, and comment on the readability of the items (Knafl et al., 2007). If a student's reasoning did not match their answer choice or was unclear to the researcher, additional probing questions were asked to clarify their interpretation of the item."	Specific. "All interviews were transcribed and coded for significant statements and emergent themes, based on each item and its corresponding scale (Creswell, 2013). The strategy for coding centered on readability and the degree of consistency among participants' interpretations of the items. If students repeatedly report dissimilar interpretations of an item, then there can be no consensus on what the score of that item really means. This is problematic, as it negatively affects the validity of the scale and the inferences that can be drawn from the scores."	Specific. "Split the larger interest scales into two smaller subscales, which resulted in improved fit of the CFA models. Our revisions to the self-efficacy and effort beliefs scales were guided more by the way students interpreted the items. When participants assign varied interpretations to a particular item, the meaning of that item is obscured and integrity of the score and construct associated with that item is compromised. Items that were found to illicit ambiguous or frequently incongruent responses from participants were flagged as problematic. Those items that also demonstrated lack of fit or redundancy in the CFA models were removed from the corresponding scale. Overall, our interview results for the items retained in each scale suggest that students consistently and adequately understood the meaning of the items."
Taskin, V.; Bernholt, S.; Parchmann, I. An Inventory for Measuring Student Teachers' Knowledge of Chemical Representations: Design, Validation, and Psychometric Analysis. Chem. Educ. Res. Pr. 2015, 16 (3), 460–477.	Chemical Representations Inventory (CRI)	Developed new	none reported	General. "Think-aloud interviews with university students were conducted in order to again improve the comprehensibility of the items by excluding or changing misleading wordings."	None reported.	None reported.

Galloway, K.R.; Bretz, S.L. Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. J. Chem. Educ. 2015, 92 (7), 1149–1158.	Meaningful Learning in the Laboratory Instrument (MLLI)	Developed new	13	None reported.	None reported.	General. No modifications needed.
Cooper, M.M.; Kouyoumdjian, H.; Underwood, S.M. Investigating Students' Reasoning about Acid–Base Reactions. J. Chem. Educ. 2016, 93 (10), 1703-1712.	Assessment tasks for the study	Developed new	5	None reported.	None reported.	General. No modifications needed, as "[they] found that students were consistently able to provide explanations that were aligned with the provided prompts, and therefore no further interviews were necessary."
Harshman, J.; Yezierski, E. Characterizing High School Chemistry Teachers{'} Use of Assessment Data via Latent Class Analysis. Chem. Educ. Res. Pr. 2016, 17 (2), 296–308. https://doi.org/10.1039/C5RP00215J. ; Harshman, J.; Yezierski, E. Test–Retest Reliability of the Adaptive Chemistry Assessment Survey for Teachers: Measurement Error and Alternatives to Correlation. J. Chem. Educ. 2016, 93 (2), 239-247.	Adaptive Chemistry Assessment Survey for Teachers (ACAST)	Developed new	14	None reported.	None reported.	Specific. Full list of modifications provided in appendix.
Harsh, J. A. Designing Performance-Based Measures to Assess the Scientific Thinking Skills of Chemistry Undergraduate Researchers. Chem. Educ. Res. Pr. 2016, 17 (4), 808–817.	Performance Assessment of Undergraduate Research Experiences (PURE) instrument	Developed new	7	Specific. Semi-structured and focused on any student difficulties in responding to the tasks as well as their interest in the situating topics and the overall test.	None reported.	None reported.

Hale, L. V. A.; Lutter, J. C.; Shultz, G. V. The Development of a Tool for Measuring Graduate Students¦' Topic Specific Pedagogical Content Knowledge of Thin Layer Chromatography. Chem. Educ. Res. Pr. 2016, 17 (4), 700–710. https://doi.org/10.1039/C5RP00190K.	Instrument to measure content knowledge and pedagogical content knowledge of thin layer chromatography for graduate teaching assistants	Developed new	5	Specific. “Cognitive interviews were performed to ascertain the validity of the results by determining whether what was conceptualized as PCK was actually measured by the test items. For example, whether a participant's response was coded for PCK but arose from a process of reasoning rather than from their knowledge of students (Willis et al., 1999). During the interviews, participants were asked to report how they answered questions related to PCK using a protocol based on a “verbal probing technique” to provide a sense of the cognitive processes that participants engaged in when responding (Willis et al., 1999). Interview questions included general probing questions such as “How did you arrive at that answer?” and more specific questions such as “Do you recall when you first realized that students would find that challenging?””	Specific. “Cognitive interviews were transcribed verbatim and coded by one of the authors for the sources of knowledge invoked to respond based on a coding scheme adapted from Hill (Hill, 2008). The Hill coding scheme was modified for the question type (open questions only), the topic specific PCK model used to conceptualize the test questions, and the sources of knowledge that are unique to graduate students (i.e. research as well as teaching) (Table 4). Hills' original code for test-taking skills was not used because the PCK portion of the instrument did not include multiple choice or forced response type questions on which respondents might be likely to apply test-taking skills. The code for “mathematical reasoning” was adapted to “chemical reasoning” and divided between reasoning that originated from knowledge gained through coursework and knowledge gained through research, which is an important distinction between graduate students and pre-service teachers”	Specific item revisions described.
Dick-Perez, M.; Luxford, C. J.; Windus, T. L.; Holme, T. A Quantum Chemistry Concept Inventory for Physical Chemistry Classes. J. Chem. Educ. 2016, 93 (4), 605–612.	Quantum Chemistry Concept Inventory (QCCI)	Developed new	56	None reported.	None reported.	General. Results from the preliminary instrument prompted some minor language adjustments and the addition of two items relating to spectroscopy.
Vrabec, M.; Prokša, M. Identifying Misconceptions Related to Chemical Bonding Concepts in the Slovak School System Using the Bonding Representations Inventory as a Diagnostic Tool. J. Chem. Educ. 2016, 93 (8), 1364–1370.	Bond Representations Inventory (BRI)	Modified pre-existing	28	General. “After the reviews of our experts, we administrated the BRI in one class of 28 students aged 15−16. After we graded the test, we conducted interviews with students regarding their remarks about the test.”	Specific. “The students complained the most about the question No. 16: most of the students have never met with a similar representation of bonds as in our tasks. In our school system, the students are not introduced to this kind of bond representation until organic chemistry. Students also reacted to the term octet rule, which is commonly used but it is not included in the new NChES. In the current chemistry textbook for the first high school grade, this term is included in expanded learning material.”	Specific. “Taking into account the findings that came from the prestudy, we partially edited some formulations in the Slovak translation. In this case, we followed the principle that the nature and the character of the given item should not be different from those of the original. The changes we made were mostly related to the Slovak mutation, and based on the above stated facts we concluded that it is possible to use BRI in a wider drafted study.”

Versprille, A.; Zabih, A.; Holme, T.A.; McKenzie, L.; Mahaffy, P.; Martin, B.; Towns, M. Assessing Student Knowledge of Chemistry and Climate Science Concepts Associated with Climate Change: Resources To Inform Teaching and Learning. J. Chem. Educ. 2017, 94 (4), 407-417.	Chemistry and Climate Science Concepts Associated with Climate Change Assessment	Developed new	2 per general chemistry course (total not given)	General. Think-aloud interviews where students were asked to respond to questions and describe their thoughts out loud.	None reported.	None reported.
Liu, Y.; Ferrell, B.; Barbera, J.; Lewis, J. E. Development and Evaluation of a Chemistry-Specific Version of the Academic Motivation Scale (AMS-Chemistry). Chem. Educ. Res. Pr. 2017, 18 (1), 191–213. https://doi.org/10.1039/C6RP00200E.	Academic Motivation Scale (AMS-Chemistry)	Modified pre-existing	11	Specific. Students were asked to complete the AMS-Chemistry instrument, consisting of 28 items. "Upon completion of the instrument, the students read each item aloud and explained their reasoning for the answer choice they made. If a student's reasoning did not match their answer choice, probing questions were asked in order to clarify their interpretation of the item and how it matched their answer choice and reasoning…In addition to asking probing questions regarding a single item and its interpretation, clarity was sought when a student's response to an item did not match their responses to the other items in the same subscale (e.g., to experience). As the instrument contains four items per subscale, each item should be measuring similar aspects of student motivation and thus elicit similar responses."	Specific. "Transcripts from eleven interviews were reviewed and coded to produce results informing the readability, response consistency, and interpretation of the AMS-Chemistry items. "	Specific. "Two items (13 and 14, shown below) required wording changes to address subscale consistency."

Undersander, M. A.; Lund, T. J.; Langdon, L. S.; Stains, M. Probing the Question Order Effect While Developing a Chemistry Concept Inventory. Chem. Educ. Res. Pr. 2017, 18 (1), 45–54.	Question prompt designed for the study	Developed new	7	General. Interviews were conducted using a think-aloud protocol.	None reported.	Specific. Two versions created based on interview findings.
Bretz, S.L.; Mayo, A.V.M. Development of the Flame Test Concept Inventory: Measuring Student Thinking about Atomic Emission. J. Chem. Educ. 2018, 95 (1), 17-27.	Flame Test Concept Inventory	Developed new	Two rounds: 7 and 3	Specific. "The validation interviews followed a think-aloud protocol where students re-answered all items, providing information about how well they understood the items, and which strategies (elimination, guessing) they used to choose an answer."	Specific. "If a student was not guessing and chose a distractor, this indicated the distractor was appealing and/or rooted in the alternative conceptions identified in the four-phase flame test interviews."	Specific. Deletion of two items, the addition of one item, and revisions in wording to others in the first round. "No additional modifications were deemed necessary" following the second round with 3 high school students.
Connor, M. C.; Shultz, G. V. Teaching Assistants' Topic-Specific Pedagogical Content Knowledge in 1H NMR Spectroscopy. Chem. Educ. Res. Pr. 2018, 19 (3), 653–669.	NMR questionnaire made for the study	Developed new	2	General. Interviews were semi-structured.	None reported.	Specific. Detail removing a question based on interview data.

Shah, L.; Rodriguez, C. A.; Bartoli, M.; Rushton, G. T. Analysing the Impact of a Discussion-Oriented Curriculum on First-Year General Chemistry Students' Conceptions of Relative Acidity. Chem. Educ. Res. Pr. 2018, 19 (2), 543–557.	Acidity Concept Inventory (ACIDI)	Modified pre-existing	19	Specific. "Students' cognitive interview responses to the statement "read the question and tell me what you thought it was asking" were used to further establish content validity (see Appendix). These types of interviews are valuable for assessing the nuances of particular assessment items For each item, students were asked to (i) explain what they thought the question was asking, (ii) why they chose the response they did, and (iii) why they did not choose the remaining choices."	None reported.	None reported.
Hansen, C.; Barbera, J. Assessing Affective Differences between a Virtual General Chemistry Experiment and a Similar Hands-On Experiment. J. Chem. Educ. 2019, 96 (10), 2097-2108.	Chemistry Laboratory Anxiety Instrument (CLAI); Attitude Toward the Subject of Chemistry (ASCI); Virtual and Physical Experimentation Questionnaire (VPEQ)	Modified pre-existing	10	Specific. "During the interview, participants completed the items from the anxiety, emotional satisfaction, intellectual accessibility, usefulness of lab, equipment usability, open-endedness of lab, initial interest-feeling, and initial interest-value scales, one item at a time. For each item, they were asked to provide their reasoning for selecting a particular response to determine if there was evidence for response process validity."	Specific. "If the rationale students provided was not aligned with the researcher's intended meaning, the item was flagged for modification or removal. If students gave a rationale that was aligned with the intended meaning, there was validity evidence that the item was being interpreted as intended and could be used in future implementations of the scale in the given context."	Specific. "All 10 students struggled to correctly interpret the item, 'There is opportunity for me to pursue my own experimental interest' from the VPEQ usefulness of lab scale. All students interviewed were confused about what the term 'experimental interests' meant. Therefore, this item was discarded, and all other items were retained."

Abell, T.N.; Bretz, S.B. Development of the Enthalpy and Entropy in Dissolution and Precipitation Inventory. J. Chem. Educ. 2019, 96 (9), 1804-1812.	Enthalpy and Entropy in Dissolution Concept Inventory (E2DCI)	Developed new	7 and 9	None reported.	General. Informal analysis to determine if respondents were "properly interpreting the questions and that they were selecting the right answer for the right reasons."	Specific. Additional option added to Item 14 after the response process interviews because several students indicated that they were picking the correct answer only on the basis of the representation of the aqueous solution while ignoring the representation of the solid.
Trate, J. M.; Fisher, V.; Blecking, A.; Geissinger, P.; Murphy, K. L. Response Process Validity Studies of the Scale Literacy Skills Test. J. Chem. Educ. 2019, 96 (7), 1351–1358.	Scale Literacy Skills Test (SLST)	Modified pre-existing	58	Specific. "Students...were instructed to verbalize their problem-solving process as they worked (either on paper or on a calculator) to arrive at an answer and were instructed to verbalize anything written down in the work space along with their selected answer. The interviewer again asked follow-up questions as needed to clarify the student's process."	Specific. Interviews transcribed and coded based on process used by respondent for specific items. Independent raters identified items that posed potential threats to response process. Discrepancies in coding discussed until agreement achieved.	Specific. "Response process data collected during these studies identified five items in Chemistry and four items in Anatomy and Physiology that are threats to the validity of the scale literacy skills test. Two out of the five items identified in Chemistry that were changed prior to the response process study in Anatomy and Physiology were revealed to no longer pose a threat in the new discipline. Two of the remaining three items were written in such a way that changing the stem or distractors was not likely to eliminate the threat and were anticipated to remain a threat in the new discipline."

(1) Roche Allred, Z. D.; Bretz, S. L. Development of the Quantization and Probability Representations Inventory as a Measure of Students' Understandings of Particulate and Symbolic Representations of Electron Structure. J. Chem. Educ. 2019, 96 (8), 1558–1570.	Quantization and Probability Representations Inventory (QuPRI)	Developed new	18	Specific. "Students were purposefully sampled from all the students who volunteered to participate in the study, primarily on the basis of their total score on the QuPRI in order to allow the authors to investigate whether all students, regardless of their total score, were interpreting the items and distractors as intended. During the interviews, students were given a blank copy of the QuPRI and asked to answer each item while thinking aloud about how they interpreted the item and each response option."	None reported.	Specific. "Response process validation interviews resulted in the addition of one new item (for a total of 27 items) and minor wording modifications to two other items. Two items were modified on the basis of students' comments that they found distractors to be confusing during the responses process interviews. For example, item 6 in the pilot version of the QuPRI asked students to choose the best description for an interpretation of the statement saying there is 90% probability of finding an electron within a three-dimensional representation depicting an orbital. About one-fourth of the students who participated in the response process validity interviews found one of the distractors 'contradictory', subsequently leading them to examine each of the options using test-taking strategies."
Lazenby, K.; Rupp, C. A.; Brandriet, A.; Mauger-Sonnek, K.; Becker, N. M. Undergraduate Chemistry Students' Conceptualization of Models in General Chemistry. J. Chem. Educ. 2019, 96 (3), 455–468. Lazenby, K.; Stricker, A.; Brandriet, A.; Rupp, C. A.; Becker, N. M. Undergraduate Chemistry Students' Epistemic Criteria for Scientific Models. J. Chem. Educ. 2020, 97 (1), 16–26.	Models in Chemistry Survey (MCS)	Developed new	8	Specific. Maximum variation sampling to invite diversity of responses from pilot administration. Students were asked to complete the tasks in the MCS again, this time describing their thought process aloud as they did so.	Specific. Analysis focused on construct-underrepresentation (i.e., not measuring the full range of the intended construct) and construct-irrelevance (i.e., measuring constructs outside the scope of the intended construct).	General. "On the basis of this analysis, we modified, expanded, or eliminated several survey questions."
Lutter, J. C.; Hale, L.+A103:F103 V. A.; Shultz, G. V. Unpacking Graduate Students' Knowledge for Teaching Solution Chemistry Concepts. Chem. Educ. Res. Pr. 2019, 20 (1), 258–269.	Questionnaire to measure pedagogical content knowledge of solution chemistry concepts	Developed new	7 and 2	Specific. "The interviews included items from the questionnaire and items that identified the graduate students' attitude towards teaching, prior teaching experience, and academic background. Because participants were not observed, these items might not differentiate between PCK and simple reasoning strategies (Willis et al., 1999). Therefore, "verbal probing techniques" outlined by Willis were incorporated to unearth why the interviewee answered survey items the way they did. The interviewer exercised substantial latitude, in particular when following lines of inquiry regarding the	None reported.	Specific. No modifications needed. "All participants described a clear understanding of the wording and intent of each question and no particular questions were identified as problematic during interviews. The interviews did reveal differences in the reported source of knowledge in answering each PCK question. Participants indicated that their general teaching style as well as their specific knowledge for teaching solution chemistry topics was attributed to their own experiences as students, their experiences as TAs, or through a reasoning process."

				sources of knowledge used when responding to the questions."		
Hosbein, K. N.; Barbera, J. Development and Evaluation of Novel Science and Chemistry Identity Measures. Chem. Educ. Res. Pract. 2020, 21 (3), 852–877.	Interest, mastery experience, and verbal persuasion scales	Developed new	8	Specific. Verbal-probing. Survey given to respondents and asked to read each item aloud; respondents asked to describe what items are asking, which option they would choose, and why they chose the option they did.	Specific. Data analyzed for evidence of readability and interpretation; alignment of item response and students' explanation provided evidence of RP validity.	General. No modifications needed.
Hosbein, K. N.; Alvarez-bell, R.; Callis-duehl, K. L.; Sampson, V.; Wolf, S. F.; Walker, J. P. Development of the Investigation Design, Explanation, and Argument Assessment for General Chemistry I Laboratory. 2020.	Investigation Design, Explanation, and Argument Assessment for General Chemistry I Laboratory (IDEAA-GC1)	Developed new	none reported	None reported.	None reported.	General. Questions made more explicit, but no details provided.
Atkinson, M. B.; Popova, M.; Croisant, M.; Reed, D. J.; Bretz, S. L. Development of the Reaction Coordinate Diagram Inventory: Measuring Student Thinking and Confidence. J. Chem. Educ. 2020, 97 (7), 1841–1851.	Reaction Coordinate Diagram Inventory (RCDI)	Developed new	10	Specific. Think-aloud interviews in which students given a copy of test, read each item aloud, explain what the item was asking, choose an answer, and explain why they chose that answer.	None reported.	Specific. Changes based on (1) students missing an option on Q7 and (2) adding an option to Q17 based on one respondents Incorrect reasoning. Additional interviews (n = 14) conducted; no discussion of protocol, analysis, or modifications for this second round.

Stephenson, N. S.; Duffy, E. M.; Day, E. L.; Padilla, K.; Herrington, D. G.; Cooper, M. M.; Carmel, J. H. Development and Validation of Scientific Practices Assessment Tasks for the General Chemistry Laboratory. J. Chem. Educ. 2020, 97 (4), 884–893.	Instrument created for study	Developed new	5	Specific. Concurrent think-aloud followed by retrospective. Students asked to talk about their reasoning as they responded to prompts (Livescribe). Post-task, students asked about readability/clarity	General. "Student feedback responses were summarized immediately following the interviews, and were used to improve the readability and clarity of the prompts"	General. Modifications made, but no details provided.
Vincent-Ruz, P.; Meyer, T.; Roe, S.G.; Schunn, C.D. Short-Term and Long-Term Effects of POGIL in a Large-Enrollment General Chemistry Course J. Chem. Educ. 2020, 97 (5), 1228-1238.	Instrument created for study	Developed new (modified items from established scales)	5	Specific. "These interviews consisted of think alouds in which approximately five male and female students from diverse backgrounds read each item, explained in their own words what the item was asking, and explained the reasoning behind their elected response to the item."	None reported.	None reported.
Komperda, R.; Hosbein, K. N.; Phillips, M. M.; Barbera, J. Investigation of Evidence for the Internal Structure of a Modified Science Motivation Questionnaire II (MSMQ II): A Failed Attempt to Improve Instrument Functioning across Course{,} Subject{,} and Wording Variants. Chem. Educ. Res. Pr. 2020, 21 (3), 893–907.	Modified Science Motivation Questionnaire II (MSMQ II)	Modified pre-existing	40	Specific. "During the interview, students were provided a paper copy of the SMQ II and randomly presented with either the science or chemistry wording of the items. Students read all of the items silently and circled their responses on the original frequency-based scale. Students were asked to explain their responses to a subset of items, which were identified by the research team as having potentially good or poor fits to the response scale and/or a hypothesized item category based on a prior study (Komperda et al., 2018a). Next, students were asked to go over the entire survey again and explain if any item responses would change if the other wording (science or chemistry) were substituted."	Specific. "The audio files of the interviews were transcribed by a commercial transcription service. All analysis was conducted on the interview transcripts, led by author Hosbein. Cognitive interview analysis (Beatty, 2004) was used by author Hosbein to establish a coding rubric. This entailed carefully reading each transcript and documenting the language students used in explaining their responses. These reviews were used to establish initial codes regarding students' scale-based language. Example codes are provided in Table S4 (ESI†). All coding was completed using QSR NVivo 8. Author Hosbein and a secondary (non-author) researcher independently coded all 40 transcripts using the initial rubric. After discussing discrepancies in coding, the rubric was revised and additional transcripts were independently re-coded. The final rubric had an acceptable Cohen's kappa value of 0.8 (Hallgren, 2012) and was used by author Hosbein to re-code the remaining transcripts."	Specific. "An additional version of the item 'The science I learn is relevant to the world around me' was added to the survey in order to test whether more general wording could avoid having the item align with the career items. The second item, I3, was revised into two different wordings to address different reasons for why science or a specific discipline may be 'meaningful' to students, which is by increasing their understanding or appreciation. The phrase 'world around me' was used in these revisions to align with the revision to item I1. Additionally, students describing their response to item G4 emphasized thinking about their grade in terms of worrying about requirements for graduate or professional school so an additional item was added to explore the 'worry' aspect of thinking about grades."

Smith, K. C.; Alonso, V. Measuring Student Engagement in the Undergraduate General Chemistry Laboratory. Chem. Educ. Res. Pr. 2020, 21 (1), 399–411.	survey developed to investigate students' engagement during general chemistry laboratory sessions	Developed new	11	Specific. "Eleven general chemistry students were interviewed individually to gauge their understanding of the various survey items. Each student was interviewed on either nine or ten of the survey items, with each survey item being addressed by either three or four students. In each student's interview, each survey item presented to them was printed out on a slip of paper and presented to them one at a time. Each student was asked to read aloud each survey item, and to describe in their own words what the survey item meant to them."	None reported.	Specific. "Several small changes were made to the wording of some of the survey items. One such change involved the words 'excited' and 'intrigued' in some of the survey items; students indicated that such words were too enthusiastic, and that the word 'interested' would be more fitting, which reinforced the importance of such item wording."
Wang, Y.; Rocabado, G.A.; Lewis, J.E.; Lewis, S.E. Prompts to Promote Success: Evaluating Utility Value and Growth Mindset Interventions on General Chemistry Students' Attitude and Academic Performance J. Chem. Educ. 2021, 98 (5), 1476-1488.	Attitude Toward the Subject of Chemistry (ASCI); Attitude Toward the Subject of Chemistry—Utility and Emotion (ASCI-UE)	Modified pre-existing	None reported	None reported.	None reported.	General. "Refined adjectives to improve clarity based on student responses during cognitive interviews."

Atkinson, M.B.; Bretz, S.L. Measuring Changes in Undergraduate Chemistry Students' Reasoning with Reaction Coordinate Diagrams: A Longitudinal, Multi-institution Study. J. Chem. Educ. 2021, 98 (4), 1064-1076.	Reaction Coordinate Diagram Inventory (RCDI)	Modified pre-existing	10	Specific. Each validation interview lasted approximately 45 min and followed a think-aloud protocol, where each student was asked to (1) read each item on the RCDI aloud, (2) explain in their own words what the item was asking, (3) choose an answer to each item and explain why they chose that answer, (4) rank the confidence of their response to each item, and (5) explain why they did not choose the other available answer options.	None reported.	General. No modifications needed.
Lazenby, K.; Becker, N. M. Evaluation of the Students' Understanding of Models in Science (SUMS) for Use in Undergraduate Chemistry. Chem. Educ. Res. Pr. 2021, 22 (1), 62–76.	Students' Understanding of Models in Science (SUMS)	Modified pre-existing	11	Specific. "During the interviews, students were asked to respond to each of the items on the SUMS, describing their thought process aloud as they did so. Students were also prompted to comment on statements or language that they found confusing or vague."	Specific. "Video recordings of the response process interview were used to identify items which participants identified as confusing or vague; we also analyzed the video to identify instances where students' response processes indicated that the item elicited construct-irrelevant ideas (i.e., measuring constructs outside the scope of the intended construct)."	Specific. Removed items 8, 15, 16, and 23
Connor, M. C.; Glass, B. H.; Shultz, G. V. Development of the NMR Lexical Representational Competence (NMR-LRC) Instrument As a Formative Assessment of Lexical Ability in 1 H NMR Spectroscopy . J. Chem. Educ. 2021.	NMR Lexical Representational Competence Instrument	Developed new	5	Specific. Cognitive interviews serve as a common means of collecting evidence of response process validity. During each interview, participants verbalized their thinking involved in responding to each item. Participant recruitment and data collection continued until participants expressed no new thinking, indicating data saturation had been achieved.	General. These verbalizations allowed for the identification of multi- ple-choice questions with wording that resulted in participants either responding correctly without knowing given terms or responding incorrectly while knowing given terms.	Specific. The wording was then altered as necessary to help ensure that questions would be interpreted as intended. Alterations are described in the Supporting Information.

ALTERNATIVE METHODS TO INVESTIGATE RESPONSE PROCESS

- *Open-response questions*, which prompt respondents to provide their thoughts, opinions, beliefs, *etc.* about the test-taking experience. This method also allows for greater number of participants if administered in the form of an online survey, for example.[1–3] For example, test developers could administer a pilot version of a test and provide a space for respondents to comment on their perspectives, item clarity, *etc.* in writing following each item.
- *Vignettes*, which are short stories or descriptions of a hypothetical respondent used to investigate a respondent's cognitive processing.[4–6] Vignettes can be useful if researchers are interested in evaluating whether participants respond to the complete item or focus only on specific elements of the item. For example, the vignette "Maria's chemistry exam is coming up and they only have a few days to study. How should Maria study for the exam in this short amount of time?" could be used to provide insight into how respondents' approach studying and their beliefs about effective study strategies.
- *Focus groups* can be used to facilitate dialogue between respondents, allowing respondents to discuss feelings, thoughts opinions, *etc.* through social interaction. Focus groups can be useful for exploring topics not previously considered by the test developers and/or collecting a variety of respondent perspectives in a single session.[7–9] For example, after individually completing a pilot version of a test, respondents could be invited to a focus group where they discuss the relevance of items and/or how they engaged with specific items.
- *Eye-tracking* can be used to study where participants look, in what order, and the amount of time they spend looking various elements of an item. These studies provide insight into indirect cues to attention and cognitive process when responding to items, allowing researchers to make inferences about psychological processes and cognitive operations. Eye-tracking has also been used to collect data on participants' pupil dilation, item fixation and duration alongside

think-aloud interviews to evaluate challenges during response process.[10–14] An example of when eye-tracking may be useful in the context of response process could be if a test item involves complex images/figures/symbols (*e.g.*, graphs, molecules, reactions)—here, eye-tracking could be used to determine how frequently and/or long respondents attend to these parts of the items.

LITERATURE REVIEW METHOD

We conducted a literature search in *Journal of Chemical Education (JCE)* and *Chemistry Education Research and Practice (CERP)* for any articles that used cognitive interviews or response process interviews to gather evidence for response process validity. Title, abstract, and keywords were searched using the queries summarized in the table below. Review articles were excluded from the search. We began by using the broad search terms *questionnaire*, *survey*, and *instrument,* all of which returned >200 results in both journals. Author NS reviewed titles and abstracts from these search results to identify additional query items to narrow the search. We identified *concept inventory* as a possible query to narrow down the search results. We further narrowed the results by including the term *response process* with the previous queries. These searches yielded 161 total results in *JCE* and 117 total results in *CERP,* respectively. We then reviewed the full text of each article and eliminated any that did not mention conducting cognitive interviews or gathering evidence for response process validity. Articles in which these were only mentioned in the limitations were excluded. If details were present in the supporting information or appendices then the article was included. After scrutinizing each full text, we narrowed our results down to 43 articles for our analysis. All searches did not have a bound date range (*i.e.*, all-time).

We conducted a qualitative analysis of these 43 articles using *NVivo 12.* We searched the methods, results, and discussion sections from each article and identified where the authors described cognitive interviews or response process validity. We then deductively coded these sections using the following four questions based on Tourangeau's model: (1) How do you (the respondent) interpret the item? (2) What information do you believe to be relevant for responding to the item? (3) How and why did you decide on their given response? (4) What challenges did you face when providing their response within the given item structure, if any?

We also reviewed articles for how response process validity evidence was reported: *none reported* (no response process investigations reported), *general descriptions provided* (*i.e.*, the authors mention an analysis, but do not provide explicit details), and *specific*

descriptions provided (*i.e.,* the authors explain their process and resulting decisions with details). For modifications, we captured whether the authors stated that *no modifications were needed* following response process investigations.

Author NS conducted the analysis, and author JD reviewed author NS' analysis for reliability. Any disagreements were discussed between the two authors prior to reaching consensus. For example, one article's response process protocol described how students were asked "What is the item trying to learn from you?". JD and NS initially disagreed on which stage of Tourangeau's model was being addressed by the question; however, after discussing, the authors decided that the prompt fit best under "How do you (the respondent) interpret the item?".

Journal	Query	# of hits	# of full texts reviewed	# of articles included in full analysis
Journal of Chemical Education	"concept* inventory*"	56	161	26
	Questionnaire*	4228		
	Survey*	3957		
	Instrument*	10700		
	"response* process*"	28		
	"Cognitive interview*"	3		
	Instrument* AND "response* process*"	21		
	Questionnaire* AND "response* process*"	17		
	Survey* AND "response* process*"	20		
	"concept* inventory*" AND "response* process*"	16		
Chemistry Education Research and Practice	"concept* inventory*"	43	117	17
	Questionnaire*	307		
	Survey*	390		
	Instrument*	503		
	"response* process*"	18		

	"Cognitive interview*"	14		
	Instrument* AND "response* process*"	15		
	Questionnaire* AND "response* process*"	7		
	Survey* AND "response* process*"	14		
	"concept* inventory*" AND "response* process*"	6		

ADDITIONAL FIGURES FROM LITERATURE REVIEW

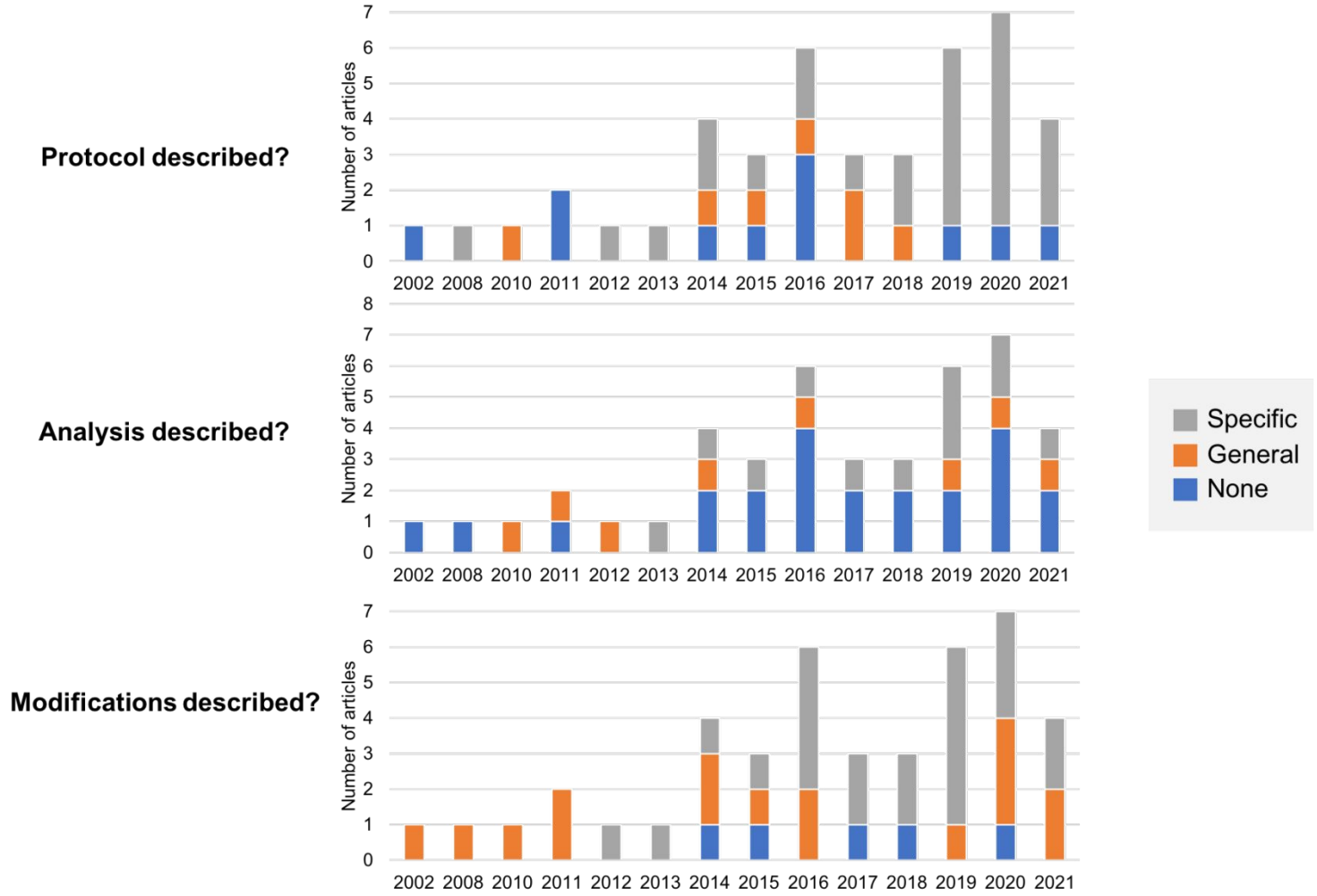


Figure 1: How response process reporting has changed over time in the CER literature.

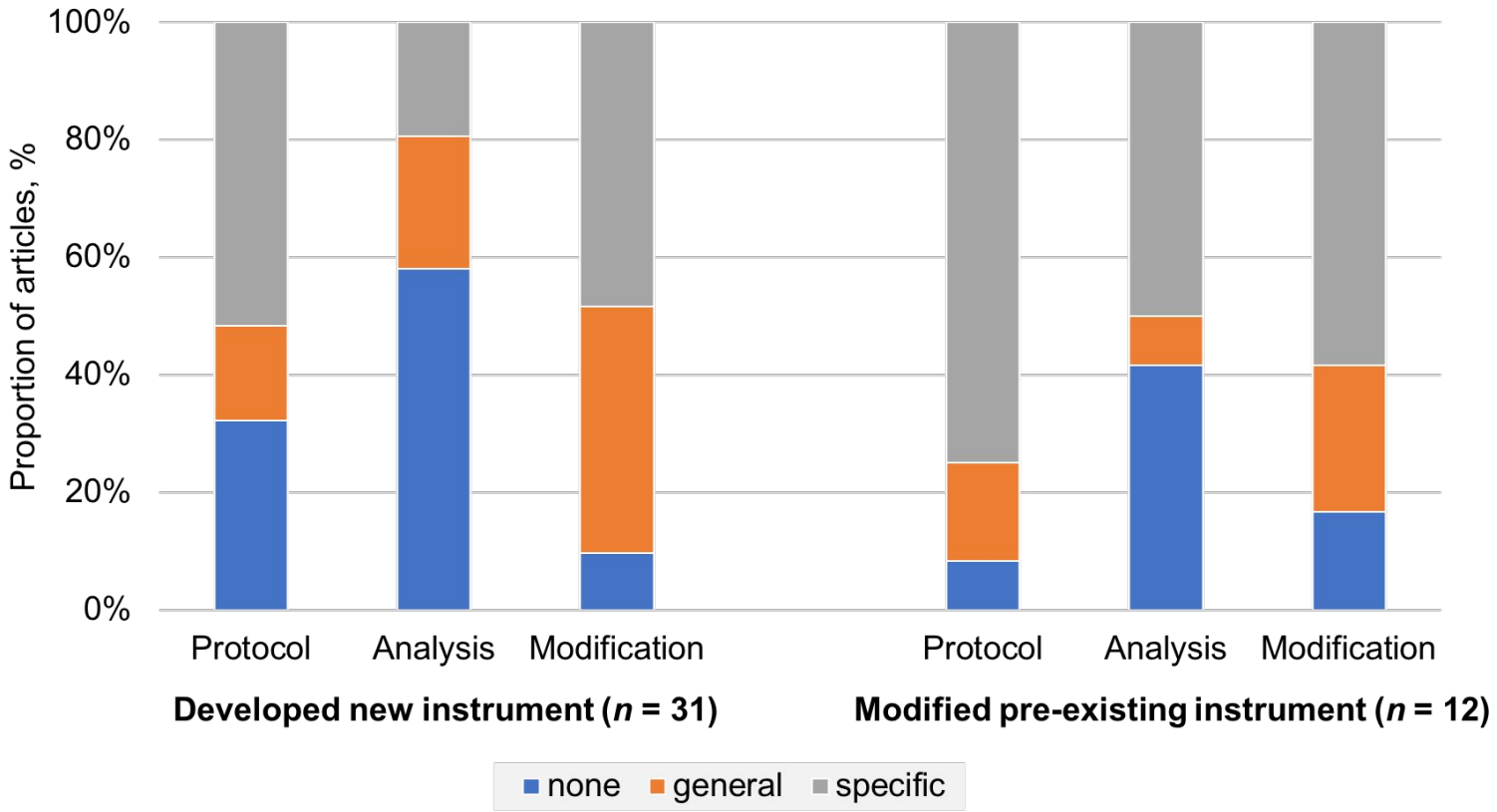


Figure 2: Reporting of response process investigations for articles that describe developing new instruments (left) and modifying pre-existing instruments (right).

CONDUCTING INTERVIEWS REMOTELY

There are many opportunities afforded by conducting interviews remotely, including: hosting instruments online, minimizing video and audio recording costs, and allowing researchers to record participants' drawings as they speak.[15] In this section, we highlight some considerations for conducting cognitive interviews remotely.

First, many online survey methods allow tests to be completed virtually (*e.g.*, SurveyMonkey, Qualtrics). Hosting tests on these online programs can be beneficial for facilitating response process interview dialogue when screen-sharing is possible. With screen-sharing, both the interviewer and the interviewee are able to view the relevant items simultaneously, similar to in-person.

Videoconferencing software with built-in recording capabilities also make it easier to manage audio and video data from interviews. For example, Zoom can be used to simultaneously record audio and video during interviews, which may allow researchers to dedicate more focus to interviewees' statements, notetaking, and formulating questions in more dynamic, think-aloud approaches. Recording audio and video simultaneously may be more cost-effective than using separate devices (*i.e.*, video camera, microphone).[15] Other useful options include the ability to transcribe audio for cloud recordings on Zoom, which can streamline interview analyses.[16] However, care must be taken with regards to research ethics in these cases, as cloud storage of interview data may conflict with existing institutional review board (IRB) principles and approvals. Hosting the discussion virtually may also be beneficial in reducing interviewees' feelings of pressure, which may afford more direct articulations of participants' response processes.[15]

There are also many options available for drawing molecules in an online setting, critical to many areas of chemistry. Participants can draw using pen and paper with either a webcam pointing to the page or by taking a photo/scan and sending to the researcher (*e.g.*, email, upload to a secure folder, drop into the videoconference chat). They could draw using a notepad app (*e.g.*, Drawboard, Notability).[17,18] Whiteboard apps offer a combination of options, including drawing, photo sharing, and collaboration (*e.g.*, Miro, Mural).[19,20] With all the electronic approaches, participants can readily share screen with the interviewer. If participants are working collaboratively, the interviewer can mute and turn their own video off to fade into the background. Offering a few different options provides opportunities for students with variable access to technologies to participate equitably.

(1) Cooper, M. M.; Sandi-Urena, S.; Stevens, R. Reliable Multi Method Assessment of Metacognition Use in Chemistry Problem Solving. *Chem. Educ. Res. Pract.* **2008**, *9* (1), 18–24.
(2) Lacosta-Gabari, I.; Fernández-Manzanal, R.; Sánchez-González, D. Designing, Testing, and Validating an Attitudinal Survey on an Environmental Topic: A Groundwater Pollution Survey Instrument for Secondary School Students. *J. Chem. Educ.* **2009**, *86* (9), 1099–1103.
(3) Russell, L. B.; Hubley, A. M. Some Thoughts on Gathering Response Processes Validity Evidence in the Context of Online Measurement and the Digital Revolution. In *Understanding and Investigating Response Processes in Validation Research*; Zumbo, B. D., Hubley, A. M., Eds.; Springer International Publishing: Vancouver, B.C., 2017; pp 229–249.
(4) Jonas, K. G.; Markon, K. E. Modeling Response Style Using Vignettes and Person-Specific Item Response Theory. *Applied Psychological Measurement*. 2019, pp 3–17.
(5) Stecher, B.; Le, V. N.; Hamilton, L.; Ryan, G.; Robyn, A.; Lockwood, J. R. Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educ. Eval. Policy Anal.* **2006**, *28* (2), 101–130.
(6) Brovelli, D.; Bölsterli, K.; Rehm, M.; Wilhelm, M. Using Vignette Testing to Measure Student Science Teachers' Professional Competencies. *Am. J. Educ. Res.* **2014**, *2* (7), 555–558.
(7) Snijkers, G.; Haraldsen, G.; Jones, J.; Willimack, D. *Designing and Conducting Business Surveys*; Wiley: Hoboken, NJ, 2013.
(8) Snijkers, G. Cognitive Laboratory Experiences: On Pre-Testing Computerised Questionnaires and Data Quality, Utrecht University, 2002, Vol. PH.D. Thes.
(9) Jobe, J. B. Cognitive Psychology and Self-Reports: Models and Methods. *Qual. Life Res.* **2003**, *12* (3), 219–227.
(10) Neuert, C. E.; Lenzner, T. A Comparison of Two Cognitive Pretesting Techniques Supported by Eye Tracking. *Soc. Sci. Comput. Rev.* **2016**, *34* (5), 582–596.
(11) Neuert, C. E. How Effective Are Eye-Tracking Data in Identifying Problematic Questions? *Soc. Sci. Comput. Rev.* **2020**, *38* (6), 793–802.
(12) Lenzner, T.; Kaczmirek, L.; Galesic, M. Seeing through the Eyes of the Respondent: An Eye-Tracking Study on Survey Question Comprehension. *Int. J. Public Opin. Res.* **2011**, *23* (3), 361–373.
(13) Menold, N.; Kaczmirek, L.; Lenzner, T.; Neusar, A. How Do Respondents Attend to Verbal Labels in Rating Scales? *Field methods* **2014**, *26* (1), 21–39.
(14) Elling, S.; Lentz, L.; de Jong, M. Users' Abilities to Review Web Site Pages. *J. Bus. Tech. Commun.* **2012**, *26* (2), 171–201.
(15) Trate, J. M.; Teichert, M. A.; Murphy, K. L.; Srinivasan, S.; Luxford, C. J.; Schneider, J. L. Remote Interview Methods in Chemical Education Research. *J. Chem. Educ.* **2020**, *97* (9), 2421–2429.
(16) Zoom Video Communications Inc. Using audio transcription for cloud recordings https://support.zoom.us/hc/en-us/articles/115004794983-Using-audio-transcription-for-cloud-recordings- (accessed Jan 3, 2021).
(17) Drawboard Pty Ltd. Drawboard https://www.drawboard.com/ (accessed Jun 30, 2021).
(18) Ginger Labs. Notability https://www.gingerlabs.com/ (accessed Jun 30, 2021).

(19) RealtimeBoard Inc. Miro https://miro.com/about/ (accessed Jun 30, 2021).
(20) Tactivos Inc. Mural https://www.mural.co/ (accessed Jun 30, 2021).

13.6 Appendix 6: Chapter 9 Supporting Information

Supporting Information

DEFINING LEARNING COMMUNITIES

Learning communities can be broadly defined as a group of people that are provided a semi-regular space or structure to work together in pursuit of a shared goal.[1] Learning communities can take four generic forms:[2,3]

- **"Curricular learning communities** [that] are made up of students co-enrolled in two or more courses (…) linked by a common theme"[3]
- "**Classroom learning communities** [that] treat the classroom as the locus of community-building by featuring cooperative learning techniques and group process learning activities as integrating pedagogical approaches"[3]
- "**Residential learning communities** [that] organize on-campus living arrangements so that students taking common courses [can] live in close physical proximity",[3] boosting opportunities for interactions and supplemental learning;
- **"Student-type learning communities** are communities that are specially designed for targeted groups, such as academically underprepared students, historically underrepresented students, students with disabilities, *etc.*"[3]

CHEMISTRY GRADUATE STUDENT DEIR COMMUNITIES AND INITIATIVES

The following section are examples of activities and initiatives from various CGDEIRCs across Canada and the US. The authors hope that these examples can support readers in organizing similar activities within their own contexts.

Canadians Working for Inclusivity in the Chemical Sciences, Engineering, and Technology

In Canada, the Canadians Working for Inclusivity in the Chemical Sciences, Engineering and Technology (CWIC) Network was formed in 2017 to create a resource hub between three groups interested in improving DEIR in chemistry at the University of Toronto, University of Alberta, and the University of Saskatchewan.[4,5] These groups previously worked independently to provide mentorship to a range of like-minded groups of individuals locally, but members of individual chapters volunteer to run the Network in order to connect individual

chapters at institutions across Canada. The number of CWIC chapters has quickly grown from three (2017) to fourteen (2020) and the Network now spans across the entire country.

The Network highlights achievements and events of chapters, including via social media (*e.g.*, Twitter), providing guidance to new chapters with step-by-step student resource group formation instructions and access to a shared online community space (*e.g.,* Slack) including at least one student from each CWIC chapter for general advice. Chapters can be created at any institution and are inclusive to students, post-doctoral fellows, and faculty of any and all identities. Members, who voluntary engage with CWIC chapters and activities, are welcome to engage as much or as little as they prefer. The Network also liaises with the Chemical Institute of Canada through their in-house DEIR committee, WIDE (Working for Inclusivity, Diversity, and Equity), to help guide policy decisions related to the Canadian chemistry community. CWIC organizes various national events for students, including online social events, book clubs, and workshops/seminars on DEIR in chemistry and managing mental health in graduate school. Two of CWIC's major initiatives will be discussed in the following subsections: the Leaders Overcoming Great Inequalities in Chemistry Conference and the My STEM Stories science communication workshop.

Leaders Overcoming Great Inequalities in Chemistry (LOGIC) Conference

The CWIC Network organizes the annual Leaders Overcoming Great Inequalities in Chemistry (LOGIC) conference. The conference is open to all members of the Chemical Institute of Canada (CIC). Early LOGIC conferences were organized by individual chapters (4-8 members) as a two-day event and has grown as the Network has grown. The majority of participants in LOGIC 2020 were chemistry graduate students (>50%), with the remaining attendees being professors, research associates, postdoctoral researchers, and industry professionals. On average, the majority of attendees are female (88%) with a smaller percentage male (11%) and nonbinary (1%). In order to reach a larger audience, this conference is held as a satellite event to the Canadian Chemistry Conference and Exhibition (CCCE), which gives the opportunity for graduate students who are already travelling to this major conference to also attend LOGIC without additional travel expenses. Organizations that sponsor the conference

also receive complimentary tickets, bringing in industry perspectives. Sponsors who choose not to use the ticket are encouraged to sponsor the ticket price of a student, which is awarded via an application process. The conference provides attendees with expert-led DEIR seminars, networking opportunities, workshops, flash talks, and poster sessions. LOGIC is organized by volunteers from individual CWIC chapters, strengthening ties between local chapters across the entire Network.

Historically, the majority of LOGIC attendees have been members of equity-deserving groups in chemistry, and so LOGIC has generally been a space that supports and focuses on the experiences of equity-deserving groups in chemistry. The diversity of attendees allows for conversations from different perspectives and allows students to seek mentorship and career advice from role models in more advanced positions. Some participants are personally invited by the LOGIC organizing committee to ensure that attendees come from different career paths and career levels to support undergraduate and graduate students looking for mentors.

The conference also provides attendees with an opportunity to contribute to working groups. In these groups, participants have the option to share their thoughts on various topics, including mental health in graduate school, intersectionality in STEM, LGBTQ2S+ experiences in STEM, transitioning from academia to industry, and others. The actionable suggestions from these group discussions are published on the CWIC Network website as a preliminary resource for other groups to engage in similar, sustained discussions and enact change. These conversations can be held at smaller, one-off events where organizers can lead multiple round-table discussions per session. Lastly, flash talks allow attendees to present their scientific research alongside their contributions to DEIR, providing attendees from equity-deserving groups to discuss their work and collaborate with others within the community.

My STEM Stories

My STEM Stories is a science communication workshop organized by CWIC that focuses on equipping equity-deserving groups in STEM with storytelling tools to highlight their diverse career paths. The event is free to attend (due to funding from Natural Sciences and Engineering Research Council of Canada) and attracts participants from chemistry, engineering, biology,

health science, social science, ecology, environmental science, and math/statistics. Similar events may include a fee if funding is not available to organizers.

Attendees attend virtual workshops hosted by industry professionals in scientific storytelling, science communication, video production and editing (with freely available software), and content accessibility. With these skills, attendees are then tasked with creating a video describing their scientific work and journeys as a scientist from an equity-deserving group in STEM. One major focus of the event is to direct attention to accessibility when creating online content, such as the importance of including subtitles in your videos, considering color contrast, *etc.* Workshop videos are published on YouTube as an open education resource available for anyone who wishes to provide students in K-12 and undergraduate science programs with diverse examples of STEM professionals.

University of Ottawa CWIC Chapter

The University of Ottawa (uOttawa) CWIC chapter was formed in the summer of 2020 and aims to create a community of individuals focused on celebrating the diversity within the uOttawa Chemistry and Biomolecular Sciences department and to minimize the barriers faced by equity-deserving groups in chemistry within its local context. Participation is voluntary and general funding is provided through the CWIC Network and uOttawa's Campus Vibez (club funding). Initiatives include social events and workshops/seminars on DEIR in chemistry and managing mental health in graduate school. Other initiatives to be discussed in the following sections include a DEIR-focused book club, social media campaigns to highlight chemists from equity-deserving groups, and a graduate student mentorship program.

DEIR-focused book clubs

In collaboration with other departmental student associations (Biology, Earth and Environmental Sciences, *etc.*) the University of Ottawa CWIC chapter contributes to a student-led, DEIR-focused book club for members of uOttawa's Faculty of Science. The book club aims to connect students, staff, and faculty and start conversations related to the challenges faced by equity-deserving groups in both STEM and the general population.

Since the beginning of the book club, five different novels and four different podcast episodes have been discussed, including "The Skin We're In" by Desmond Cole, "The

Inconvenient Indian" by Thomas King, "The Autobiography of a Transgender Scientist" by Ben Barres, and "Academic Ableism" by Jay Timothy Dolmage.[6–9] Meetings are organized once per week, with two people selected as the "hosts" to direct discussion for the meeting on a select set of chapters. Participants discuss different perspectives about each book/podcast, and participants are welcome to provide as much or as little insight as they prefer. New members are always welcome, and the book club facilitators explicate the importance of a safe space that welcomes the perspectives of all participants.

To provide access to the book club for as many students as possible, the uOttawa CWIC Chapter provides bursaries to reimburse book costs for students that participate in the book club and they have also created a freely accessible book library. When a book cannot be accessed easily through the public or university library system, the chapter purchases copies of the current book on an as-needed basis. Funding for these two supports is provided by the uOttawa club's association (Campus Vibez), the CWIC Network, and community donations.

Social media campaign to highlight chemists from equity-deserving groups

The uOttawa CWIC chapter runs a regular social media campaign on Twitter to highlight individuals from equity-deserving groups within the department (#InspiringPerson). The goal of this campaign is to (1) amplify the achievements of chemists from marginalized backgrounds and (2) help improve the visibility of role models within the chemistry community. Each month, the chapter posts a profile highlight of one individual in our department (with their consent) that showcases their professional/academic accomplishments in chemistry/STEM, as well as their personal achievements, which can often be considered "invisible work". Over the past year the #InspiringPerson campaign at uOttawa has been the most interacted series of tweets by the uOttawa CWIC Network Chapter.

Graduate student buddy programs

The uOttawa CWIC Chapter supports a graduate student-led mentorship program that pairs junior graduate students (and upper-year undergraduate students) with upper year graduate student mentors. The transition to graduate school can be daunting (especially for students from equity-deserving groups), and the program aims to make new students more

comfortable with the process and get quickly acquainted with peers. Long-term, the program aims to create a strong network of future students, current students, and alumni through on-going mentorship connections. In the first year of the program, author LEM participated as a mentor for new incoming student, and she has found that the program can help with feelings of imposter syndrome and supports students' mental health.

When students voluntarily sign up for the program, they are provided a questionnaire asking about their personal and professional goals. The program then matches students based on their interests for a four-month term (one semester) where upper year student mentors can help answer questions related to campus, courses, program requirements, social events, the local area, *etc.* from new student mentees. Other departmental graduate student associations at uOttawa (*e.g.*, Biology) have taken inspiration from this initiative and have created similar initiatives that include a brief workshop on effective mentorship for graduate students.

University of Minnesota-Twin Cities (UMN-TC) Graduate Student DEIR Community

Underrepresented Students in STEM Symposium-Twin Cities (USSS-TC)

The graduate student DEIR community at UMN-TC organizes a local research symposium for BIPOC students in STEM. This initiative—called the Underrepresented Students in STEM Symposium-Twin Cities, or USSS-TC—supports students from local institutions within the Twin Cities (undergraduate institutions and community colleges) that have fewer opportunities to present their work and/or attend national conferences due to monetary restrictions or lack of resources. At the symposium, participants are assigned volunteer graduate student mentors from UMN-TC to assist them in creating their presentations. The mentorship aspect of the symposium also provides students with resources related to undergraduate research experiences (URE), internship positions, and graduate school applications. The symposium includes research discussions, panel discussions on graduate school applications and graduate student life, and presentations by local graduate students.

General funding for the graduate student DEIR community's work has been provided by various sponsors, including 3M, the Center for Sustainable Nanotechnology, the Center for Sustainable Polymers, ACS Analytical Chemistry Journal, and the Midwest American Association for Clinical Chemistry.

Research Experience for Undergraduates Workshop

The graduate student DEIR community has contributed to the development of an research experience for undergraduates (REU) workshop directed at BIPOC students in the Twin Cities.[10] At the workshop, students meet with professors and graduate students from different programs at UMN to discuss graduate school, REUs, and "a day in the life" in graduate school. Students have an opportunity to hear from REU coordinators from fields that include Mathematics, Astronomy, Chemistry, Materials Research Science and Engineering Centers, National Science Foundation Centers, Plant Biology. Students also have opportunities to speak with the REU coordinators one-on-one to develop connections and learn from role models. Students are also invited to bring their CVs/resumes to get feedback from coordinators so that the former can strengthen their application based on their program(s) of interest. Approximately 30 students from local institutions, including primarily undergraduate institutions, community colleges, and tribal colleges, are invited to the workshop. Transportation and food are offered to students and funding is provided by the UMN Medical School.

University of Iowa (UI) Graduate Student DEIR Cohort

Faculty, staff, and graduate students in the Department of Chemistry at the University of Iowa (UI) recently re-established a new departmental DEIR committee. The formal committee's purpose is to identify DEIR-related issues within the department, then implement actions or resources that address these issues. The structure of monthly meetings typically involves 1) discussing potential DEIR issues from faculty, staff and students, 2) deliberating on plausible action items, and 3) providing progress updates on these items. By the end of the committee's first semester in December 2020, a list of action items was finalized, including tasks such as: ratifying and publishing a departmental anti-racism statement, contributing events to our college's spring initiative titled "The intersection of STEM and Race" (described in following sections on "Colloquium series on DEIR graduate student leadership" and "Identity awareness week"), and identifying inequities in graduate-level comprehensive exams. The committee is composed of three faculty members (one of whom is designated as chair), two staff members, three graduate students, a post-doctoral scholar, and an undergraduate student.

Membership within the committee is voluntary. The DEIR committee also regularly coordinates with other formal department committees (*e.g.*, graduate admissions, graduate education, *etc.*) to provide guidance and to help prioritize DEIR alongside other institutional committees.

In conjunction with the formation of the department's DEIR committee, chemistry graduate students at UI created a Graduate Student DEIR Cohort. This student organization was formed by and comprises a diverse group of chemistry graduate students focused on improving DEIR within the UI Department of Chemistry. Monthly meetings are structured around three topics: (1) educating students on an array of social justice themes (*e.g.,* intersectionality, racism, collective trauma, *etc.*), (2) sharing of professional development opportunities for students (undergraduate and graduate) and (3) planning and executing DEIR initiatives for the department. Initiatives conducted in the 2020/2021 are described in the below sections. Future initiatives that are in-progress include student-led anti-racist pedagogy workshops, a resource campaign for international student success, and development of a DEIR-service award that formally recognizes advocacy work for mention on a curriculum vitae. Three representative leaders, who serve as members on the DEIR committee, are selected through a yearly democratic vote to ensure diversity and equity within the organization. The student cohort regularly communicates with the faculty committee issues and needs expressed by students by scheduling monthly cohort meetings before committee meetings.

The UI DEIR Committee is provided general funds by the department, which is compiled at the beginning of the fiscal year. Funding for initiatives for the Graduate Student DEIR Cohort are included in this itemized budget.

Microaggressions TA Training

Sponsored by the Graduate Student Senate at UI, the goal of this student-led training is to equip graduate teaching assistants (GTAs, both returning and new) with a heightened sensitivity to identity and microaggressions within the classroom. The training is embedded into the traditional training offered to all TAs at the beginning of the academic year. Throughout the workshop, GTAs discuss implicit bias, types of microaggressions, how

microaggressions manifest in undergraduate classrooms, and ways they can intervene when a microaggression occurs. Participants also engage in several interactive activities:

- **Identity activity**: participants practice discussing the concept of identity and learn how implicit biases manifest around certain identities.
- **Live polls**: done after the concept of microaggressions is introduced to assess how well attendees can recognize microaggressions in case studies.
- **Breakout sessions**: done after a case study is presented, participants engage in multiple discussions and on-screen annotation activities to incorporate what they learn into their teaching

Roles that are essential to this training include breakout group facilitators and tech moderators, especially if training occurs in a virtual context.

Colloquium series on DEIR graduate student leadership

The UI DEIR Graduate Student Cohort has created a colloquium series wherein graduate students engaged in DEIR work at other institutions are invited to present their contributions to the UI Department of Chemistry. This seminar series provides students with an additional professional development opportunity related to their DEIR work. Speakers are selected based on their experience in DEIR work, such as DEIR advocacy work in chemistry. Attendance (voluntary) for the colloquia in 2021 from the department averaged 60 people for a duration of 1.5 hours. During/after the presentations, attendees can also participate in breakout discussions to brainstorm ideas for future DEIR initiatives and reflect on what these initiatives aim to address. The interactive nature of these colloquia center discussions of DEIR and equity-deserving groups in chemistry, and attendees have the opportunity to learn from peers and build networks within the breakout discussions. Provided by the DEIR Committee, honoraria (in this case, $100 USD) for the graduate student speakers are essential to encourage similar colloquium-based initiatives in the future and avoid tokenizing speakers' experiences.

Identity awareness week

This initiative serves to provide a week of DEIR-based spaces for discussion, support, and networking centered around marginalized groups in STEM, including but not restricted to: 1) LGBTQ2S+ scientists, 2) women in science, 3) scientists with visible/invisible disabilities

and mental health challenges, and 4) international scientists. Each day, a virtual discussion room—with voluntary participation by students, faculty, staff, *etc.*—is facilitated by faculty and staff who are certified to work with their respective groups (*e.g.*, LGBTQ2S+ facilitated by a professor with LGBTQ2S+ SafeZone training,[11] mental illness and disabilities facilitated by university counseling services staff member, *etc.*) and is co-facilitated by volunteers from the DEIR Graduate Student Cohort. During the event, facilitators present on issues relevant to their assigned identity topic and highlight resources and opportunities for academic enrichment through DEIR.

Through organizing and participating in these communities and initiatives, author HTD has gained significant experience in inclusive leadership, DEIR event planning, and working with diverse university personnel. He has also learned much about how to work as an active DEIR leader to help uplift the needs and voices of graduate students from equity-deserving groups, especially on a predominantly Caucasian university campus.

Outreach activities beyond institutional contexts

Engagement in DEIR communities within our institutional contexts has also allowed individuals to create opportunities to engage in outreach activities beyond our institutional roles. The following section will describe outreach initiatives that were led and organized by author RSR and other students at UMN-TC to support STEM students from equity-deserving groups in Minnesota.[12,13] These outreach activities provide an example of how graduate students can contribute to broader communities that support the development of cultural wealth within equity-deserving groups at various levels of science education.

To address the inaccessibility of science literacy to non-English speakers, RSR worked with the NSF Funded Center for Sustainable Nanotechnology (CSN) to translate scientific concepts and information into Spanish to facilitate connections between chemistry topics and everyday experiences for middle schoolers and adults who are non-English speakers. The CSN has a science communication blog, *SustainableNano*, that examines topics related to sustainability, agriculture, chemistry and nanoscience, in which RSR has translated their blog posts in Spanish in *Nanosostenible*.

RSR also organized a chemistry outreach event (*Spooky Science*) for underrepresented and underserved groups in the Twin Cities through hands-on demos and experiments. Funded by the MN ACS Chapter, a local team of volunteers led by RSR and the Society of the Advancement of Chicano/Hispanic and Native Americans in Science (SACNAS) Chapter at UMN-TC worked with a local school in the Twin Cities which has a predominantly LatinX and Black student population, with 90% of their students on free or reduced lunch. As a student of the LatinX community, a woman, and a student who also participated in the free/reduced lunch program due difficulties her parents faced, RSR felt a personal responsibility to serve the same communities she was once a part of. To best serve the school, activities were labeled in English, Spanish, and Somali to ensure that activities were inclusive and that students and parents were all able to engage effectively. Activities include scenarios that relate chemistry to everyday life and scientific explanations of phenomena. Questions were crafted such that any student from grades K-8 can engage with the content. Students are also provided a "bingo card" that was stamped each time they answered a question at an activity station. Completing the bingo card awarded students a prize.

Author RSR's motivation for planning outreach activities and events for younger BIPOC students stems from the lack of similar opportunities they had growing up. Having to navigate higher education as first-generation student from a low-income, immigrant family, her experiences have allowed her to identify gaps that exist within institutional support for students. She strives to provide younger students with the mentors and allies she never had and/or felt uncomfortable asking for due to her position in the existing culture.

EXPANDED LIST OF OPPORTUNITIES FOR FACULTY SUPPORT

- Sharing CGDEIRC work/initiatives with students/colleagues *via* email, social media, lectures, office hours (*e.g.*, posting event posters on office door), and meetings (*e.g.*, research group meetings)
- Inviting students to faculty meetings to engage in active dialogue on issues/work related to DEIR, especially if these items impact students. Acknowledging power dynamics and ensuring that students' voices are safe and valued in these conversations is essential for meaningful dialogue to occur.

- Sharing funding opportunities with CGDEIRCs. Opportunities to collaborate with CGDEIRCs may also become available if funding is directed towards faculty/staff initiatives.
- Participating in CGDEIRC work/initiatives. Participation may include: engaging in workshops, serving as a faculty mentor/contact for the community, co-organizing initiatives with students, *etc*.
- Conductng research on the impacts of CGDEIRCs on students, faculty, local communities. Examples may include: qualitative research focused on experiences of community members, quantitative research on impact of CGDEIRCs on student success retention, *etc*.

REFERENCES

(1) West, R. E.; Williams, G. S. "I Don't Think That Word Means What You Think It Means": A Proposed Framework for Defining Learning Communities. *Educ. Technol. Res. Dev.* **2017**, *65* (6), 1569–1582.
(2) Zhao, C.-M.; Kuh, G. D. Adding Value: Learning Communities and Student Engagement. *Res. High. Educ.* **2004**, *45* (2), 115–138.
(3) Lenning, O.; Ebbers, L. The Powerful Potential of Learning Communities: Improving Education for the Future. *ASHE-ERIC High. Educ. Rep.* **1999**, *26* (6).
(4) Chiorean, S.; Oakley, M. S. Starting Grassroots Initiatives to Foster Inclusivity in the Chemistry Department at the University of Alberta. *Can. J. Chem.* **2021**, Accepted.
(5) Canadians Working for Inclusivity in Chemical Sciences Engineering and Technology. Canadians Working for Inclusivity in Chemical Sciences, Engineering, and Technology https://cwicnetwork.com/ (accessed Apr 1, 2021).
(6) Cole, D. The Skin We're In. Doubleday Canada 2020, pp 1–510.
(7) King, T. *The Inconvenient Indian: A Curious Account of Native People in North America*; Anchor Canada, 2013.
(8) Barres, B. *The Autobiography of a Transgender Scientist*; The MIT Press, 2018.
(9) Dolmage, J. T. *Academic Ableism*; University of Michigan Press, 2017.
(10) University of Minnesota. Rebeca Rodriguez conducts workshop for nontraditional undergraduates https://cse.umn.edu/chem/news/rebeca-rodriguez-conducts-workshop-nontraditional-undergraduates (accessed Apr 1, 2021).
(11) The Safe Zone Project. The Safe Zone Project https://thesafezoneproject.com/ (accessed Apr 20, 2021).
(12) Rodriguez, R. S. Celebrating Native American/ Indigenous History Heritage Month with notable STEMists https://sustainable-nano.com/2020/11/25/native-american-heritage-month/.
(13) Rodriguez, R. S. Notable Black STEMists: Celebrating Black History Month on Twitter.

13.7 Appendix 7: Chapter 10 Supporting Information

SUPPLEMENTAL MATERIAL

SURVEY QUESTIONS IN ORIGINAL FORM

National Graduate Student Finances Survey
The Ottawa Science Policy Network is launching a "Graduate Student Finances Survey" in order to better understand student stipends.

Access this survey in french here: https://forms.gle/6rLkLjcVhmy4ogSWA

*Required

Background

We aim to better understand graduate student finances. Student stipends are not guaranteed, and often students must work additional teaching assistant positions to obtain them. These stipends often do not cover the cost of living, let alone cover additional tuition costs. In addition, after many years of post-secondary education, graduate students often carry debt from their undergraduate degrees. This can be incredibly stressful for students.

Purpose

The information collected in this survey will be used to generate a comprehensive report on graduate student finances. While some information is known about the role of Tri-Council awards and recommendations for changes to these awards, we hope to understand how stipend values can impact graduate student life. In particular, with little to no change in the value of stipends in the last 15 years, how are graduate students paying for the basics, such as housing and groceries.

Our goal is to use this information as leverage following the 2021 Federal Election and to launch a campaign. We aim to report upon the current status of stipends for Master's and Ph.D. students as funded through the Tri-Councils, Canadian Institute of Health Research (CIHR), Natural Sciences and Engineering Research Council (NSERC), and Social Sciences and Humanities Research Council (SSHRC).

Intended Survey Participants & Data Collected

This survey is intended for current or recently graduated (within one year) students pursuing a master's or doctorate (professional, academic thesis-based, academic course-based) degree that is or will be conferred at a Canadian public university. Students should be 18 years of age or older.

The survey will collect information on demographics, current studies, personal finances, stipends/scholarships, barriers and enablers to stipend values and a Likert scale on satisfaction. The survey will take 15-20 minutes to complete.

Data Privacy & Declarations

Your responses will be held in confidence by the Ottawa Science Policy Network (OSPN), a student organization at the University of Ottawa. Aggregated, anonymous and de-identified data may be shared with representatives of student associations, university administrations, advocacy groups, legislatures and governments across Canada, including data subsets where appropriate. We will ensure that no individual can be identified from these summarized results. A summary of survey data will be made publicly available on the OSPN website.

View our data usage policy here: https://docs.google.com/document/d/1zNIQMOUCuNwS1JLrVvO2yQ0Knoyo_YnLl6Cb-d54y40/edit?usp=sharing

For more information about the survey, to ask any questions, or to access survey results at a later date, please contact Ottawa Science Policy Network at ottawasciencepolicynetwork@gmail.com. Any participant who wishes to withdraw data following survey completion may do so at any point by contacting us and providing their unique identifiers. We declare no financial/sponsorship conflicts that interfere with this survey.

Thank you for taking the time to contribute to our understanding of graduate student finances. To learn more, visit www.ottawasciencepolicynetwork.ca/financesurvey.

At the end of the survey, participants have the option to enter their email (in a separate form to maintain anonymity) to be entered to win one of four $50 gift cards of their choice!

Accessibility

OSPN is committed to the accessibility of this survey for all. If you have any issues or would like to receive additional accommodations for this survey. Here are some tips on how to use Google Form Accessibility features:

Screen readers - Google Forms are made accessible to screen readers if a person using a screen reader turns on Google's screen reader support. However, the form itself may have issues as described by WebAim's Google Forms issues. To turn on screen reader support for Google Forms, press Control + Alt + z (on Windows) or ⌘ + Option + z (on Mac) in your form. Unless you turn off this setting, screen reader support stays on whenever you use Docs, Sheets, Slides, Forms, or Drawings on the web. Take a look at the Google support page for more information on how to edit a form with a screen reader.

Keyboard shortcuts - Google Forms supports keyboard shortcuts as well. Open a list of shortcuts any time by pressing Ctrl + / (Windows, Chrome OS) or Command + / (Mac). Take a look at the Google support page for a complete list of the different keyboard shortcut options.

I consent to the use of my (aggregated and de-identified) responses to help inform efforts to better financially support Canadian graduate students *

- Yes
- No

Unique Identifier

The following questions will be used to determine "unique responses" to our survey. This will help us to ensure that we don't receive multiple responses from the same person.

- Last three digits of your postal code:
- Last three digits of your phone number

Background Information

Are you a current graduate student, at a Canadian University? *
- Yes
- No
- Recently graduated (within the last year)

What best describes your current level of study? *
If you have recently graduated, specify your completed studies and please respond to the rest of the survey based on your finances from the last year during which you pursued graduate studies.
- Professional Master's
- Research Stream Master's
- Course Based Master's
- Professional Doctoral (ex. EdD, PharmD, DBA)
- Research Stream Doctoral (PhD)
- Course Based Doctoral
- MD PhD
- Professional Law Degree (ex. J.D.)
- Professional Business Degree (ex. MBA)
- Other:

What year of study are you in? *
- First Year
- Second Year
- Third Year
- Fourth Year
- Fifth Year or More
- Graduated

What is your age? *
Note: Participation in the survey is for students 18 years of age, or older.
- 18 - 23
- 24 - 27
- 28 - 33
- 34 - 40
- 40 +
- Prefer not to answer

Do you identify as a person with a disability? *
Why are we asking you to self identify? (https://www.chairs-chaires.gc.ca/forms-formulaires/self_identification_preview-eng.pdf). The Accessible Canada Act defines disability as “any impairment, including a physical, mental,intellectual, cognitive, learning, communication or sensory impairment—or a functional limitation—whether permanent, temporary or episodic in nature, or evident or not, that, in interaction with a barrier, hinders a person’s full and equal participation in society.” (SOURCE: https://laws-lois.justice.gc.ca/eng/acts/A-0.6/)
- Yes

- No
- Prefer not to answer

Do you identify as Indigenous (First Nations, Métis or Inuit)? *
- Yes
- No
- Prefer not to answer

Select the option(s) you identify with:
All categories were taken as described from Statistics Canada. Select all that apply and use the other option to add any that were missed.
- White
- Chinese
- South Asian/East Indian (including: Indian from India; Bangladeshi; Pakistani; East Indian from Guyana, Trinidad, East Africa; etc.)
- Black
- Filipino
- Non-White Latin American (including: indigenous persons from Central and South America, etc.)
- Southeast Asian (including: Burmese; Cambodian; Laotian; Thai; Vietnamese; etc.)
- Non-White West Asian, North African or Arab (including: Egyptian; Libyan; Lebanese; Iranian; etc.)
- Japanese
- Korean
- Prefer not to answer
- Other:

Select the option(s) that you identify with: *
https://isotl.sites.olt.ubc.ca/files/2019/03/eio-gender_guidelines.pdf
- Agender
- Woman
- Gender-Fluid
- Man
- Non-Binary
- Two-Spirit
- Prefer not to answer
- Other:

Are you an International Student? *
- Yes
- No

What university are you currently attending? *
- If your institution was not listed above, please list here:

What province/territory are you currently living in? *

What field are you pursuing your degree in? *
- Life Sciences
- Medical Sciences

- Physical Sciences (Engineering, Physics, Mathematics, etc)
- Computer Science
- Social Sciences
- Arts and Humanities
- Law
- Business
- Other:

Are you pursing your degree full-time or part-time?
- Full-time
- Part-time

Faculty Affiliation (ex. Faculty of Science, Faculty of Medicine): *

Secondary Faculty Affiliation (if applicable):

Department Affiliation (ex. if you are in the Biology Department or Department of Biology, write Biology here): *

Secondary Department Affiliation (if applicable):

What best describes your program: *
- Research based only
- Research based but with course requirements
- Course based only

Personal Finances

Here, we are looking for insight into your personal finances. Take a moment to pull up your personal records, tuition reports or any other things that may assist in the following responses. We will be asking about the cost of your tuition and university payments, as well as personal details such as the cost of groceries and rent.

Please list all your responses in Canadian Dollars.

What is your ANNUAL tuition fees and other mandatory (or ancillary) university fees (before scholarships)? Please make as accurate as possible. *
Mandatory (or ancillary) fees can include payment for Bus Pass, Gym/Recreational Cost, GSA payments, Student Services, Insurance, etc. For international students, this can also include international differential fees and the international health insurance costs.

How much are you paying out of pocket for tuition (after any applicable scholarships)? *

What is your monthly cost of housing? *

Which best describes where you currently live:
See here for descriptions if needed https://askanydifference.com/difference-between-rural-and-urban-and-suburban/
- Rural
- Urban

- Suburban
- Other:

Do you rent or own your current residence? *
- Rent
- Own
- I don't have stable housing
- None of the above
- Other:

Do you currently live with: *
- Commonlaw Partner or Spouse
- Roommate(s)
- Your parents
- Alone
- An extended family member
- Prefer not to answer
- Other:

How many dependents do you have? *
- No dependents
- 1
- 2
- 3
- 4 or more
- Prefer not to answer

Do you receive financial support from any of the following: (Select all that apply) *
- Financial Need Based Scholarships
- External/Internal Scholarships
- Parents or a family member
- Provincial Student Assistant Programs
- Bank Loan
- Partner/Spouse
- Savings
- None of the above
- Other:

Do you currently work a job (or get paid for a service) outside of your graduate studies? *
This includes any money that is made in addition to your stipend. This can include Teaching Assistant only if the payment is not included in your stipend value.
- Yes
- No

If you work another job(s), how many hours a week do you work on average? *
- <5 hrs
- 5-10 hrs
- 10-20 hrs
- 20-30 hrs
- 30 hrs +

- I do not have another job

How much do you spend on average per month on: *

- Not Applicable
- $0 - $100
- $100 - $200
- $200- $300
- $300 - $400
- $400 - $500
- $500 or more

1. Groceries/Food
2. Social Activities
3. Medical (Dental, prescriptions, doctor, eye care, etc)
4. Child Care
5. Transportation Costs (Bus pass or vehicle expenses)
6. Additional Bills (Internet, Phone, Hydro, Water)

What statement best describes your financial situation? *

- I am very comfortable financially. I have enough money that I do not need to worry about monthly expenses and spend as I would like.
- I have enough. My finances are bit tight but I live within my means and can afford to provide for myself.
- Every month is tight, but I am getting by. I am often making sacrifices to pay for necessities.
- I am struggling financially. I often do not have enough to make ends meet.

What is the current value of your outstanding student debt (this includes federal or provincial student assistance, bank loans and credit cards)?

How frequently do you: *

- Never
- Occasionally
- Sometimes
- Often
- Always

1. Worry about your ability to pay bills
2. Worry about your ability to pay back your student debt?
3. Worry about your future financial security
4. Worry about your savings or ability to pay for emergency expenses
5. Worry about your ability to obtain a future job in your chosen field

How much savings or funds do you have available to be used in case of emergencies? *

- 0 to 3 months of living expenses
- 3 to 6 months of living expenses
- 6 to 12 months of living expenses

- More than 12 months of living expenses
- Prefer not to answer

Have you experienced any of the following during your graduate studies (check all that apply): *
- Stress/anxiety about finances
- Difficulty budgeting finances
- Concern about your ability to pay rent
- Concern about your ability to pay for groceries/food
- Considered leaving your studies due to concern over finances
- N/A

Would you like to add any additional information about your personal finances?

Stipend Details

WHAT IS A STIPEND?

A stipend is "a fixed amount of money provided to people who pursue unpaid work. It helps offset expenses, such as housing and food. Unlike a salary, a stipend is not a payment for hours worked or services provided."

Most graduate students pursuing a research focused program receive a stipend in order to compensate them for their work. This is provided by either the students supervisor or department, depending on the university and program. Some students are required to TA in order to receive their stipend. In these cases the TA-ship is included as part of the students funding package.

In general, the value of stipends varies greatly between universities and faculties. We developed this survey to better understand how much students are getting paid to provide insight into this variation.

Help us learn more about your stipend!

Do you currently receive a stipend (payment for your graduate studies)? *
- Yes
- No

Stipend Details

WHAT IS A STIPEND?

A stipend is "a fixed amount of money provided to people who pursue unpaid work. It helps offset expenses, such as housing and food. Unlike a salary, a stipend is not a payment for hours worked or services provided."

Most graduate students pursuing a research focused program receive a stipend in order to compensate them for their work. This is provided by either the students supervisor or department, depending on the university and program. Some students are required to TA (teaching assistant) in order to receive their stipend. In these cases the TA-ship is included as part of the students funding package.

In general, the value of stipends varies greatly between universities and faculties. We developed this survey to better understand how much students are getting paid to provide insight into this variation.

Help us learn more about your stipend!

What is the current value of your stipend per academic year (September to August; for example $17,500/year)? *
Please try to be as specific as possible.

What are the source(s) of your current stipend payment? *

- Federal Award (Tri-Councils - NSERC, CIHR, SSHRC, etc)
- Provincial Award (OGS, BCGS, etc)
- Department/Faculty
- Supervisor
- Government Institution (Health Canada, National Research Council, etc)
- Industrial Institution
- Charity/Non-Profit/Foundation
- Teaching Assistant
- Other:

How long have you been funded through this source? *

- Less than 6 months
- 6 months to less than 1 year
- 1 to 2 years
- 3 to 4 years
- 5 or more years

If your stipend comes from scholarship(s), please list the name and value:

Have you ever experienced any of the following "errors" regarding stipend payments: *
An "error" = a disruption in the regular scheduled payment of your stipend. Questions based on a study conducted the Dalhousie Working for Inclusion in Chemical Sciences Group: https://twitter.com/Dal_WIC/status/1425411679463038977

- Stipend payment was received late
- Paid too little (ie. received less than your actual stipend)
- Paid too much (ie. received more than your actual stipend)
- Stipend payment was put towards a university expense/fee without my prior knowledge
- I have not experienced any errors in stipend payment
- Other:

How many times have these "errors" in stipend payments occurred? *

- I have not experienced any errors in stipend payments
- 1 error
- 2 errors
- 3 errors
- 4 or more errors

Have you had increases or decreases in your stipend value over the course of your graduate studies? *

- My stipend has increased in value
- My stipend has decreased in value
- My stipend has both increased AND decreased in value at separate times
- My stipend value has not changed

If your stipend value has changed over the course of your graduate studies, please specify the reason or add more detail:

If you have ever received a scholarship, did you receive any of the following additional incentives OR changes after receiving your scholarship? *

- A scholarship meant to fully cover your tuition costs
- A scholarship partially covers your tuition
- A "top-up" in your stipend (ie. you recieved the scholarship PLUS an additional amount from supervisor/department)
- A decrease the total amount of take-home money
- None of the above
- N/A
- Other:

Have you previously been funded through any of these sources (not including your current stipend)? *

- Federal Award (Tri-Councils - NSERC, CIHR, SSHRC, etc)
- Provincial Award (OGS, BCGS, etc)
- Department/Faculty
- Supervisor
- Government Institution (Health Canada, National Research Council, etc)
- Industrial Institution
- Charity/Non-Profit/Foundation
- N/A
- Other:

Are you happy with the value of your stipend? (based on the Likert scale on satisfaction) *

- Very Unhappy
- 1
- 2
- 3
- 4
- 5
- Very Happy

What is your ideal yearly stipend value? *
If you could name your price, what would your yearly stipend value be?

Are there any conditions of your stipend? *

- Required to work as a Teaching Assistant/Research Assistant
- Limitations on external employment
- Maintain a grade point average/percentage
- Participate in outreach activities

- Maintain "Progress" as deemed by an advisory committee
- N/A
- Other:

If you are required to work as a Teaching Assistant/Research Assistant, how many hours a week do you commit to this? *
If you have multiple positions, please include the total hours.

- 5 hours or less
- 5 to 10 hours
- 10 to 15 hours
- 15 to 20 hours
- 20 to 30 hours
- 30 to 40 hours
- 40 + hours
- N/A

Would you like to add any additional information about your stipend?

Have you ever been asked to pay for an academic expense (conference, equipment, etc) up front during your graduate studies? *

- Yes, with the expectation it would be reimbursed
- Yes, with the expectation that it would come out of pocket
- No.
- Other:

If you have had to apply for reimbursement, how long did it take to be reimbursed? *
If you have experienced this multiple times, please provide an average.

- It has never been reimbursed
- Less than a week
- 1-4 weeks
- 2-3 months
- 3-6 months
- 6 months - 1 year
- Over a year
- N/A

Has waiting for reimbursements ever impacted your ability to pay for living expenses? *

- Not Applicable
- Yes
- No

Additional Financial Details

Describe more about why you do not receive a stipend. Are there other opportunities to receive funding? *

Do you feel like your program should provide stipends for students? *

- Yes
- No
- Maybe

If so, what would your ideal yearly stipend value be?

If you could name your price, what would your yearly stipend value be?

Have you ever been asked to pay for an academic expense (conference, equipment, etc) upfront during your graduate studies? *

- Yes, with the expectation it would be reimbursed
- Yes, with the expectation that it would come out of pocket
- No
- Other:

If you have had to apply for reimbursement, how long did it take to be reimbursed? *
If you have multiple experiences with this, please provide an average.

- Less than a week
- 1-4 weeks
- 2-3 months
- 3-6 months
- 6 months - 1 year
- Over a year
- It has never been reimbursed

Has waiting for reimbursements ever impacted your ability to pay for living expenses? *

- Yes
- No
- N/A

Additional Comments

Is there any additional comments you would like to include in your survey? If so, please list here!

Do you think there is any aspect of graduate student finances, stipends or more that we missed and you think you should be included in this study?

How did you hear about our survey?

- Social media - OSPN page
- Social media - other page
- From your graduate student association
- Through a friend/another student

Would you like to be entered into a draw for 4 x $50 gift cards?

If you would like to be entered into our draw, please go to the following link to submit your email:

https://forms.gle/ZdnFqX2ykeYQ8KBt8

NOTE: This is not required, and is completely voluntary. Please note that this information will not accompany the data previously provided in this survey, and we will not connect your email to the answers provided in this survey.

DEFINITIONS

10-hour rule: a stipulation provided by most institutions across Canada in which graduate students are eligible to work a maximum 10 hours per week to supplement their stipend

Agender: a person who identifies as not having a gender identity.

Dependent: an individual who relies on another person (i.e. family member) as a primary source of income or financial support.

Disability: Under the Accessible Canada Act, disability means an impairment or functional limitation that reduces someone's full involvement in society because of the barriers they face. They can include physical, mental, intellectual, cognitive, learning, communication and/or sensory.

Diversity: differences in the lived experiences and perspectives of people that may include race, ethnicity, religion, ability, sex, gender identity, sexual orientation, age, class, and/or socio-economic status.

Equity: involves creating policies and opportunities that allow students, especially those from marginalized groups, to have equal access to education, research, and other programs.

Gender-fluid: a person whose gender identity is not fixed.

Gender non-conforming: people who do not behave in a way that conforms to the traditional expectations of their gender or whose gender expression does not fit neatly into a category. While many also identify as transgender, not all gender non-conforming people do.

Genderqueer: people who typically reject notions of static categories of gender and embrace a fluidity of gender identity and often, though not always, sexual orientation. People who identify as "genderqueer" may see themselves as being both male and female, neither male nor female or as falling completely outside these categories.

Graduate student: someone who has earned a bachelor's degree and is pursuing additional education and/or research (i.e. master's or doctoral studies) in a specific field of study.

Inclusion: an active and ongoing process to address inequities in power and privilege. This builds a respectful and diverse community where people are welcomed, and their contributions are valued.

Indigenous: encompasses First Nations, Métis and Inuit peoples in Canada.

Non-binary: an adjective describing a person who does not identify exclusively as a man or a woman. Non-binary people may identify as being both a man and a woman, somewhere in between, or falling completely outside these categories. While many also identify as transgender, not all non-binary people do. Non-binary can also encompass identities such as agender, bigender, genderqueer or gender-fluid.

Principal Investigator: a researcher, usually a doctor or other medical professional, who leads the clinical or basic research team and regularly monitors study progress. The principal investigator is responsible for the preparation, conduct, and administration of the research grant.

Research Assistantships: A student hired to provide services in the context of carrying out research activities. Duties and tasks may include but are not limited to the performance of clerical, laboratory, or any technical tasks, or the organization of research-related activities. Guaranteed hours may form part of a guaranteed funding package offered by a supervisor or department in addition to a stipend as defined above.

Scholarship: a form of financial aid awarded to students pursuing further education. Generally, awardees are selected based on certain criteria, such as academic merit, leadership experience, community involvement, and/or financial need.

Science policy: an area of policy concerning the effective allocation of funding and resources for conducting science and research to reach goals (e.g. technological innovation, advances in healthcare, environmental protection) that ultimately serve the public interest.

Stipend: a fixed amount of money provided to people pursuing unpaid work to help offset various living expenses. Unlike a salary, a stipend is not a payment for hours worked.

Student loan: a sum of money borrowed to finance post-secondary education or related expenses. This is designed to help students afford tuition, books, supplies, and living expenses. The interest rate may be lower than regular loan types, and the repayment schedule may be deferred while the student is still in school.

Teaching assistant (TA): a graduate student who assists a professor with instructional responsibilities. This is typically a part-time paid position that one applies for. Guaranteed hours may form part of a guaranteed funding package offered by a supervisor or department in addition to a stipend as defined above.

Trans: people whose gender identity and/or expression is different from cultural expectations based on the sex they were assigned at birth. Being transgender does not imply any specific sexual orientation. Therefore, transgender people may identify as straight, gay, lesbian, bisexual, etc.

Tri-Council Scholarships: scholarships issued by one of Canada's three major research grant agencies: the Natural Sciences & Engineering Research Council (NSERC), the Social Sciences & Humanities Research Council (SSHRC), and the Canadian Institute of Health Research (CIHR).

Tuition fees (including mandatory fees): semester charges imposed on students to attend an institution of higher education, which are required as a condition of enrollment.

Two-spirit: a modern, pan-Indian, umbrella term used by some Indigenous North Americans to describe people in their communities who fulfill a traditional third-gender (or other gender-variant) ceremonial and social role in their cultures

SUPPLEMENTARY DATA

Table S1. Demographic Information.

	Number of Respondents
Total Respondents:	**1,305**
Language of Response:	
English	1030
French	275
Level of Studies:	
Course Based Master's	109
Research Stream Master's	337
Professional Master's	107
Research Stream Doctoral (PhD)	750
Year of Study:	
First Year - Master's	261
Second Year - Master's	197
Third Year or more - Master's	68
First Year - Doctoral	174
Second Year - Doctoral	132
Third Year - Doctoral	154
Fourth Year - Doctoral	141
Fifth Year or More - Doctoral	134
Recently Graduated	36
Age of Respondents:	
18-23	198
24-27	529
28-33	412
34-40	110
40+	56
Full-time Status:	
Full-time	997
Part-time	24
No Response	10
Identify with a Disability:	
Yes	135
No	1140
Prefer not to answer	30

Identify as Indigenous:	
Yes	15
No	1279
Prefer not to answer	11
Identify as:	
White	868
Chinese	84
South Asian/East Indian	110
Black	75
Filipino	5
Non-White Latin American	57
Southeast Asian	22
Non-White West Asian	112
Japanese	9
Korean	8
Prefer not to answer	25
Other	14
Identify as:	
Agender	10
Woman	825
Gender-fluid	15
Man	428
Non-binary	37
Genderqueer	3
Trans	3
Two-spirit	1
Gender non-conforming	1
Prefer not to answer	20
Identify as an International Student:	
Yes	413
No	892

Table S2. Tuition Fees from Survey Responses in Comparisons to Statistics Canada Data.

Province	**Average Tuition Fees ($CAD)**	**Count**	**Statistics Canada 2022a**
Ontario	8,972.46	358	9,385.00
Nova Scotia	7,770.70	36	10,591.00
New Brunswick	7,385.39	23	7,272.00
British Columbia	7,374.11	39	9,994.00
Saskatchewan	7,118.29	7	4,971.00
Alberta	6,837.11	55	7,143.00
Quebec	6,290.74	338	3,582.00

Manitoba	5,880.69	18	5,573.00
Newfoundland and Labrador	5,016.26	12	3,759.00
Annual	7,518.00		7,437.00

Table S3. Sample Calculations for a domestic (in-province, on campus) thesis-based graduate student attending a U15 university in Canada.

University	Location	Tuition	Ancillary Fees*	Rent (one-bedroom) per month	Rent per year	Source	$550/month in additional fees	$500/month additional forPhD	Master's Minimum Stipend	PhD Minimum Stipend
University of Alberta	Edmonton, Alberta	$ 4,486.20	$ 2,600.34	$ 1,020.00	$12,240.00	Zumper	$ 6,600.00	$ 6,000.00	$ 25,926.54	$ 31,926.54
University of British Columbia	Vancouver, British Columbia	$ 5,301.54	$ 475.48	$ 2,500.00	$30,000.00	Zumper	$ 6,600.00	$ 6,000.00	$ 42,377.02	$ 48,377.02
University of Calgary	Calgary, Alberta	$ 5,990.64	$ 1,905.19	$ 1,610.00	$19,320.00	Zumper	$ 6,600.00	$ 6,000.00	$ 33,815.83	$ 39,815.83
Dalhousie University	Halifax, Nova Scotia	$ 9,099.00	$ 1,961.24	$ 1,800.00	$21,600.00	Zumper	$ 6,600.00	$ 6,000.00	$ 39,260.24	$ 45,260.24
Université Laval	Quebec CIty, Quebec	$ 3,187.44		$ 1,100.00	$13,200.00	Zumper	$ 6,600.00	$ 6,000.00	$ 22,987.44	$ 28,987.44
University of Manitoba	Winnipeg, Manitoba	$ 3,412.60	$ 1,599.26	$ 1,090.00	$13,080.00	Zumper	$ 6,600.00	$ 6,000.00	$ 24,691.86	$ 30,691.86
McGill University	Montreal, Quebec	$ 4,195.80	$ 2,193.70	$ 1,500.00	$18,000.00	Zumper	$ 6,600.00	$ 6,000.00	$ 30,989.50	$ 36,989.50
McMaster University	Hamilton, Ontario	$ 6,304.20	$ 2,084.82	$ 1,640.00	$19,680.00	Zumper	$ 6,600.00	$ 6,000.00	$ 34,669.02	$ 40,669.02
Universitié de Montreal	Montreal, Quebec	$ 4,195.35	$ 1,012.56	$ 1,500.00	$18,000.00	Zumper	$ 6,600.00	$ 6,000.00	$ 29,807.91	$ 35,807.91
University of Ottawa	Ottawa, Ontario	$ 6,366.60	$ 1,484.16	$ 1,720.00	$20,640.00	Zumper	$ 6,600.00	$ 6,000.00	$ 35,090.76	$ 41,090.76
Queen's University	Kingston, Ontario	$ 7,697.32	$ 1,412.78	$ 1,700.00	$20,400.00	Zumper	$ 6,600.00	$ 6,000.00	$ 36,110.10	$ 42,110.10
University of Saskatchewan	Saskatoon, Saskatchewan	$ 4,698.00	$ 1,175.97	$ 1,020.00	$12,240.00	Zumper	$ 6,600.00	$ 6,000.00	$ 24,713.97	$ 30,713.97
University of Toronto	Toronto, Ontario	$ 6,210.00	$ 1,787.28	$ 2,300.00	$27,600.00	Zumper	$ 6,600.00	$ 6,000.00	$ 42,197.28	$ 48,197.28
University of Waterloo	Waterloo, Ontario	$ 6,762.00	$ 1,652.33	$ 1,800.00	$21,600.00	Zumper	$ 6,600.00	$ 6,000.00	$ 36,614.33	$ 42,614.33
University of Western Ontario.	London, Ontario	$ 6,360.00	$ 2,728.19	$ 1,680.00	$20,160.00	Zumper	$ 6,600.00	$ 6,000.00	$ 35,848.19	$ 41,848.19
First year tuition used - continuing fees are lower.	*Ancillary fees include health and dental insurance, as well as a bus pass if included by the university.									

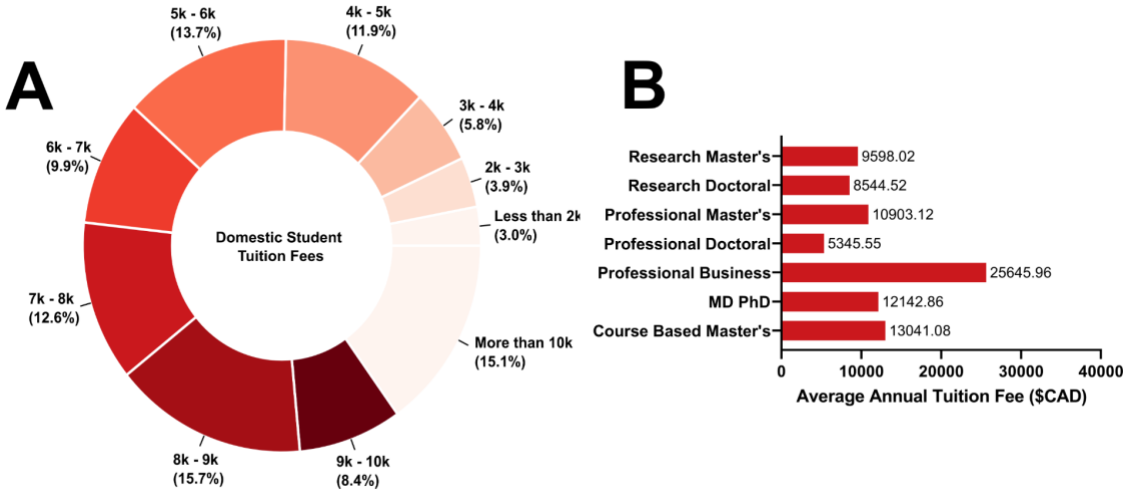


Figure S1. Tuition Fees **A** The total tuition (and other mandatory fees) paid by domestic students. Given as a percentage of students paying each amount (n=892). **B** The average tuition paid by respondents (both domestic and international) by type of degree.

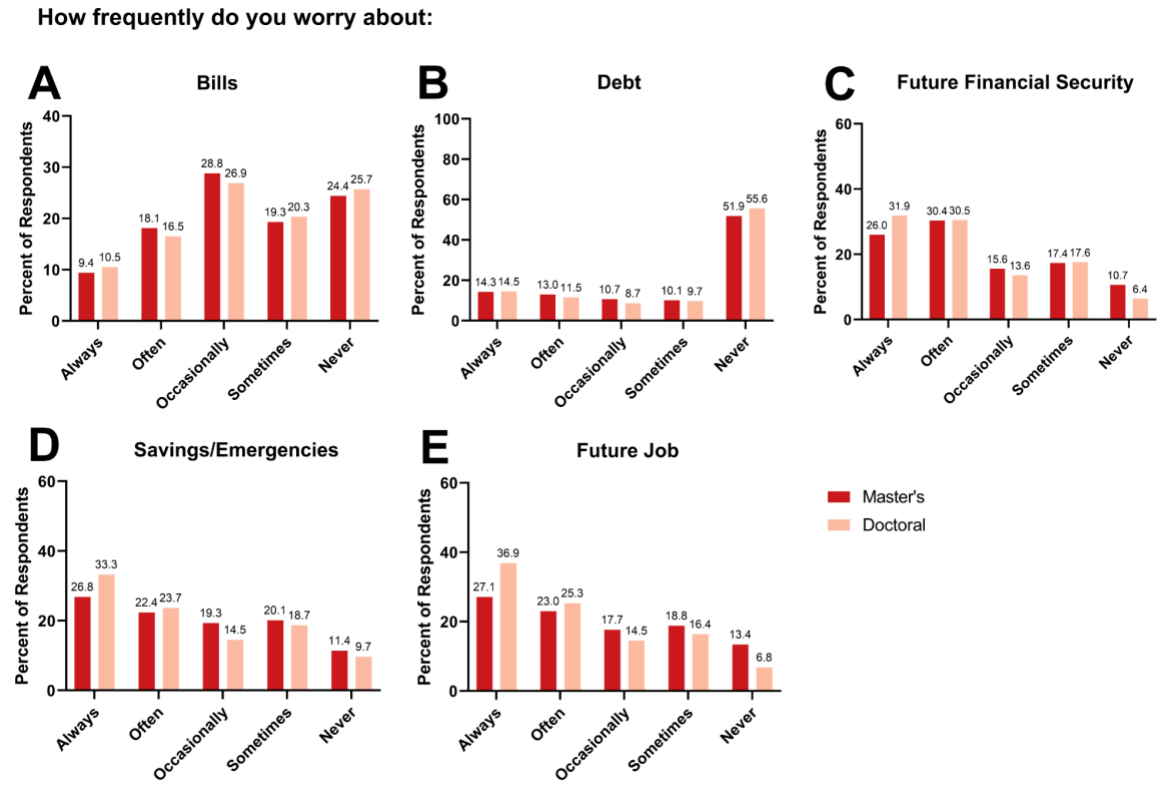


Figure S2. Financial Stresses - Master's vs. Doctoral The response for those in their Master's (n=553) vs. Doctoral (n=750) studies on how frequently they worry about their **A** ability to pay bills **B** ability to pay back their student debt **C** future financial security **D** savings/ability to pay for emergency expenses **E** ability to obtain a future job in their chosen field. Data are shown by percent in each category.

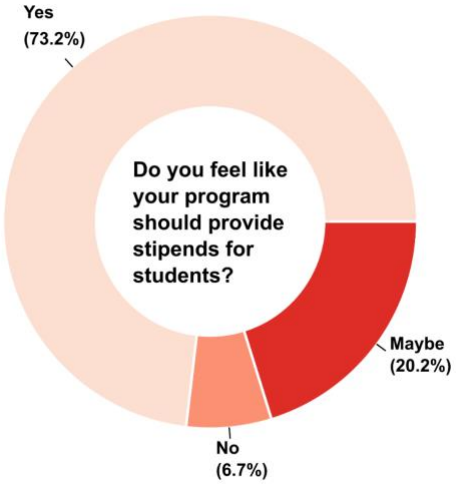


Figure S3. No Stipend The proportion of respondents who do not receive a stipend but felt that their program should (n=391).

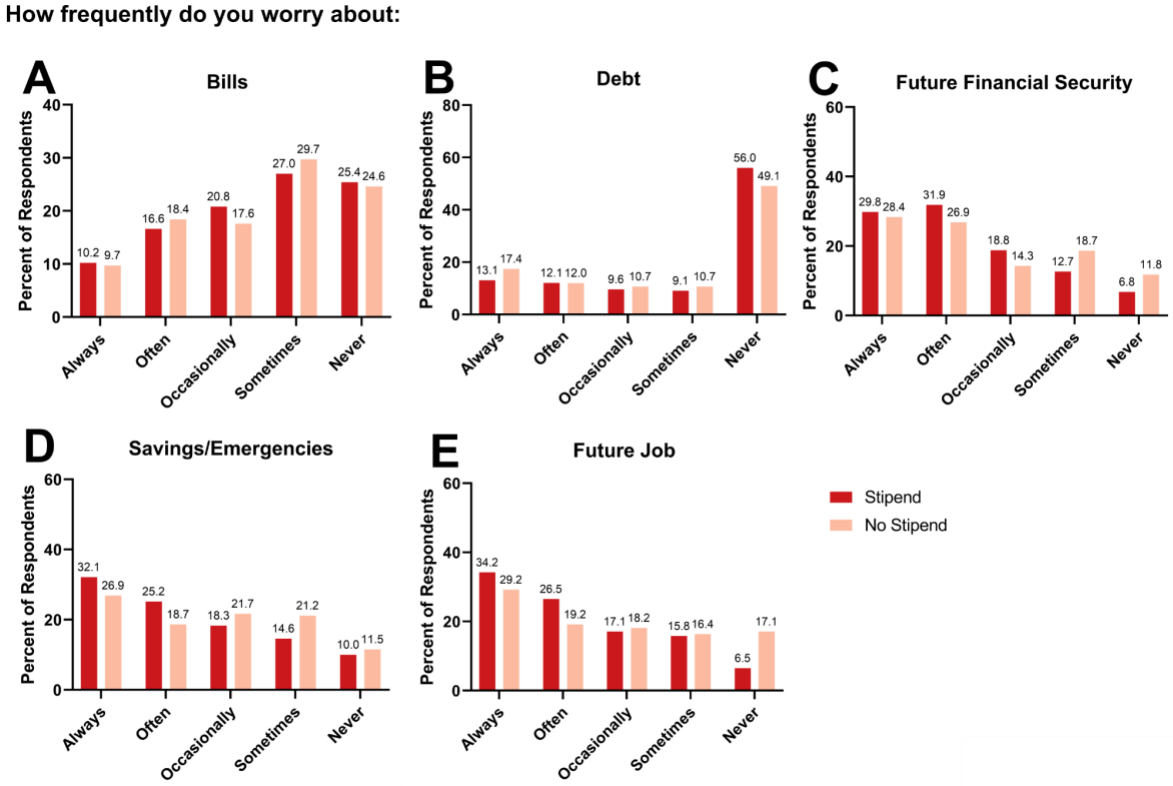


Figure S4. Stipend - Financial Stresses The response for students who received a stipend (n=914) vs. did not receive a stipend (n=391) on how frequently they worry about their **A** ability to pay bills **B** ability to pay back their student debt **C** future financial security **D** savings/ability to pay for emergency expenses **E** ability to obtain a future job in their chosen field. Data are shown by percent in each category.

13.8 Appendix 8: Chapter 11 Supporting Information

Supporting Information: Prioritizing Mentorship as Scientific Leaders

Jacky M. Deng[a,*], Salma Elgaili Ahmed[b,‡], Ernest Awoonor-Williams[c,‡], Progna Banerjee[d,‡], Magda Barecka[e,‡], Laura E. Bickerton[f,‡], Silvina A. Di Pietro[g,‡], Stanna K. Dorn[h,‡], Kevin Maik Jablonka[i,j‡], Gabriele Laudadio[k,‡], Elisabeth Kreidt[l,‡], Helena Mannochio-Russo[m,‡], Júlio Terra[n,‡], Olivia Harper Wilkins[o,‡], Saigopalakrishna S. Yerneni[p,‡], Maha Yusuf[q,‡]

AUTHOR ADDRESS: (a) Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada; (b) Centre for Advanced Imaging and The Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, Qld 4072, Australia; (c) Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States; (d) Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA; (e) 1. Department of Chemical Engineering, 360 Huntington Avenue, 02215 Boston, MA, 2. Department of Chemistry and Chemical Biology, 360 Huntington Avenue, 02215 Boston, MA; (f) EaStCHEM School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, Scotland EH9 3FJ, United Kingdom; (g) Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, United States; (h) Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; (i) Laboratory of Organic and Macromolecular Chemistry (IOMC), Friedrich Schiller University Jena, Humboldtstrasse 10, 07743 Jena, Germany; (j) Center for Energy and Environmental Chemistry Jena (CEEC Jena), Friedrich Schiller University Jena, Philosophenweg 7a, 07743 Jena, Germany (k) Department of Chemistry, Scripps Research, La Jolla, California 92037, United States; (l) Department of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Straße 6, Dortmund 44227, Germany; (m) Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA; (n) Laboratory of Sustainable and Catalytic Processing, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Station 6, Lausanne, 1015 Switzerland; (o) NASA Postdoctoral Program Fellow, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt MD 20771; (p) Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; (q) Department of Chemical Engineering, Stanford University, CA 94305, USA

Author Contributions

*Corresponding author
‡Authors contributed equally to this work.

1. Why is mentorship important?

Progna: As an early career researcher mentorship is incredibly important to me for several reasons. Mentorship is a key component of my role as an academic and is critical to creating a supportive, collaborative, and productive academic environment.

I believe that it is crucial to give back to the next generation of students and scholars by providing guidance, support, and mentorship. As an academic, I have had the privilege of working with many inspiring mentors who have helped me to develop my research skills, build my professional network, and navigate the complex academic landscape. Without their support and guidance, I would not be where I am today, and I am committed to paying it forward by serving as a mentor to students and junior colleagues. In addition to the satisfaction that comes from helping others, I believe that mentorship is also an important part of building a strong academic community. By fostering relationships with students and junior colleagues, I hope to create a network of scholars who can support each other, collaborate on research projects, and share resources and ideas. This type of collaborative community is essential for building a thriving research program and for making meaningful contributions to the academic field.

Elisabeth: Science is hard. By its very nature it is confronting you with unsolved problems and difficult tasks every day. You need a lot of (technical) knowledge, creativity and stamina to progress, and in many cases, you will lack a bit of one of these resources. This is where you will need somebody who has already progressed a bit more in their science journey, who has already solved a puzzle similar to yours or who can see the bigger picture more clearly. This is where you will need a mentor. As I have experienced it, a good mentor can make a huge contribution to how you face your challenges. Good mentorship often makes the difference between just getting it done "somehow" and extracting valuable personal growth while getting it done. While you are busy working on your everyday challenges, a mentor can help you to keep focused on the bigger picture, stay creative and keep everything in balance with your personality and personal needs. What I've learned from my mentors has become an integral part of my research personality and of how I approach science and I try to give something similar to younger chemists I am working with. Now that I am a PI of a small group, I learn that being a mentor is a complex task that needs a lot of thought and empathy. It is a challenging thing on its own (where you might need a bit of help from your own mentors) but also enormously rewarding as it allows your knowledge and abilities to contribute to more than your own work. You can make a real, sustainable impact. Apart from that the interaction with your mentees brings you fresh ideas and approaches, helping you to not become too stuck to your own established strategies and thoughts.

Ernest: Mentorship is important to me because it provides an opportunity for personal and professional growth, self reflection, and connection. Effective mentorship breeds confidence in mentees — offering guidance, support, and training as mentees navigate different aspects of their lives. Personally, I find mentorship has shaped and challenged my beliefs as a researcher and played a significant role in my scientific identity. Throughout my academic career, I have been fortunate to have great mentors who have not only provided guidance and advice but also challenged me to take ownership of difficult tasks and projects. I believe mentorship is a two-way street; it benefits the mentee as much as it benefits the mentor.

Helena: "Good" mentorship can help flourish the best of someone, make this person feel accomplished, and guide him/her toward a fruitful career. On the other hand, "bad" mentorship can be highly frustrating and negatively impact someone's vision of a specific field, for instance, even leading someone to make a dramatic change in their career. These two simple scenarios exemplify the dramatic impact that mentorship may have on someone's professional and personal life. Mentors are fundamental to guidance and encouragement throughout our scientific journey, and building a healthy and trustworthy mentor-mentee relationship is critical. The most obvious image of a mentor is the formal supervisor or advisor. However, getting advice from people of different levels who have undergone specific experiences of interest is valuable to diversifying the precious inputs someone can have. Having a comprehensive network with people from diverse fields and backgrounds is the first step toward getting a unique way of thinking.

Gabriele: Mentorship is essential in many aspects of science. It is one of the most effective ways to transfer knowledge from the mentor to the mentee, encouraging the development of both soft and hard skills. Through constant communication, the mentor and mentee create a feedback loop that connects them and allows them to share similar experiences. This connection triggers personal and professional growth for both parties.

Mentorship is important to me because it is the main way for scientists to learn and teach in a working setting. I have witnessed how the mentorship process can dramatically enhance productivity and self-awareness, leading to constant evolution that brings both the mentor and the mentee to a more mature stage. Moreover, a fruitful mentor-mentee relationship generates a healthy environment where everyone feels free to discuss, make mistakes, and provide genuine feedback in a respectful manner. In my experience, I have seen how such a relationship boosts the confidence of both parties and makes the work light, fun, and easy. I truly believe that mentorship is the only way to build a solid foundation in our scientific society, where each individual can express their personality freely.

Kevin: Mentorship is important to me because there are many things that cannot be learned from textbooks or conventional means. For example, when transitioning into a new leadership role in a lab, there are many things that are not typically taught, and having a mentor to talk through these issues can be extremely valuable. Additionally, mentors can help with exploring ideas and making career decisions. Mentors have typically faced and overcome different challenges and can provide insights into how to turn obstacles into opportunities. Mentors can also help keep momentum and motivation going, especially during challenging times. Lastly, being a mentor is important for personal growth and reflection. Mentoring can help me reflect on aspects I have not considered, and this can lead to valuable insights and growth.

Laura: I believe the majority of my successes thus far have been down to having great mentors who have facilitated my ability to work in what can be a stressful fast-paced environment. Having supportive, unbiased mentors can be the difference between giving up and thriving forward. With dyslexia and a physical impairment, I am only too aware of the adversity faced in daily life and importance of peer support. I have learnt how important it is to create a supportive working environment such that the maximum potential of each individual can be achieved, thus engendering 'in-lab' discoveries which can translate to real life applications. As a result mentorship, and additionally advocacy, are hugely important to me because I felt like these aspects were lacking throughout my school to undergraduate to PhD journey. I am sure many others have felt like they are working things out by themselves. It seems down to luck and being in the right place at the right time that I have made it this far; I believe mentoring can assist with more intentional decision making. That's why promoting STEM to the next generation with a focus on encouraging outreach and accessibility, especially for women and those underrepresented in STEM, and the promotion of chemistry to the next generation is hugely important to me.

Magda: While a typical journey of a graduate student used to be associated with a significant amount of isolation, stress and unrealistic supervisor expectations, we are now in a good position to stimulate changes that will positively influence the retention, success and the impact of researchers in the STEM field. This change does not need to come solely from the supervisors, in fact, I do advocate for catalyzing this process by graduate students. Therefore, we need to recognize the importance of cultivating the mentorship as the main pillar of academic activities and be ready to dedicate time to seeking and interacting with the mentors, and being mentors to others. I want to strongly encourage the graduate students to explore the opportunities to become mentors themselves. Becoming a mentor helps to be an informed, agile and future-ready mentee,

well positioned to understand mentors' standpoints, propose other techniques or approaches for mentorship and align expectations. I also encourage to proactively seek for various mentors, outside of our research teams or institutional structures. An excellent example of such support is the ACS Mentorship Program, which allows to connect 1-to-1 to experienced ACS members, willing to share their expertise in specific job markets (academia/industry), processes (such as transitioning into an independent research position or successful retirement), and most importantly their passion in seeing the long-term growth and success of their mentees. I believe that the choice of mentor's network can be best guided by appreciating of our own goals and boldest ambitions. Just believe in yourself!

Maha: As a beneficiary of mentoring, I strongly believe that mentors and role models can reduce barriers for underrepresented minorities (URMs) in STEM fields. Below is an example of my path to Stanford graduate school from being an international exchange student at the University of Mississippi (UM).

During my senior year of undergraduate studies in Pakistan, my hard work led to a U.S. Department of State exchange scholarship at UM, where I faced a new set of challenges as an international student and a young Muslim woman. I felt ashamed of my cultural and religious identity and was one of the few women in my engineering classes. During this tough period, I met Prof. Chen, a minority professor in ChemE, who championed me at UM. With his mentorship, I received the highest score in his thermodynamics class and initiated a research collaboration between UM and my Pakistani university, which became my Senior Honors Thesis. My UM experience exemplifies the power of diversity and mentorship in academia and the need to create an academic culture where all students feel comfortable. As a future professor, I want to strive for this goal and emulate Prof. Chen's example in effectively mentoring marginalized students.

Silvina: I am a people's person. As much as I enjoy my passion for physical sciences, I also enjoy—and feed off—social interactions. But, for me, social interaction has to serve a purpose. During my two plus decades of being a student, I learned a very simple equation: social interaction + education = mentorship. The purpose (education) was stimulated by social interaction (my mentors). This symbiotic relationship became key to appreciate and understand its importance when it comes to mentorship. Simply put, mentorship served a purpose for my growth as a professional with the aid of more experienced professionals.

At the conclusion of my dissertation defense mid-pandemic back in March of 2021, my principal investigator (PI) very eloquently concluded his congratulatory remarks with the following quote: "It takes a village". There is absolutely no way for a scientist to have a successful career if he/she was not mentored. Whether it is an unwelcome and mediocre or enriching and conductive mentorship experience, it served a purpose. Like any relationship, I find mentorship essential to my professional development. In a scientific setting, it is a perfect blend of social interaction and transferred technical knowledge.

Stanna: I would not be on the career path that I am today without phenomenal mentors that brought the best out of me and actively encouraged and helped me pursue my career goals. As

someone who comes from a rural, low-socioeconomic background and nearly failed out of organic chemistry during my undergraduate studies, I am fortunate to have had excellent mentors.

Scientific study necessitates that we remain open to new hypotheses and ways of thinking; indeed, we propose and actively explore hypotheses from a variety of angles in an attempt to obtain as full a picture as possible for a given system. It seems ironic, then, that many in the scientific community are not more invested in efforts to increase diversity, equity, and inclusion to promote a variety of perspectives on science as a whole (although this argument is often made to incentivize interest in DEI efforts, it should be noted that *everyone* should be supported and feel as though they have a place in science, regardless of whether the 'bottom line', or scientific output, is affected). While improvements have been made and are ongoing, there is a dire need for effective, inclusive mentorship in academia. Too many trainees are forced to leave their programs or feel trapped because mentors/advisers have abused the inherent power dynamic of academic settings. Mentorship is important to me because there were many times on my scientific journey that I was ready to give up and leave, but I had mentors that helped me keep going. At times, I have wondered where I would be without them. I wonder about those that do not have supportive mentors, and the institutions that tout impressive research but foster poor mentorship at the expense of trainees.

Julio: Navigating grad school and academia is not a simple task, especially for first-generation scientists/academics like many of us. We are also often not living in our home countries and not working in our mother tongues during grad school, so we end up living in a completely new world in which we have no experience. The role of a mentor to me is to provide guidance and assistance to the mentee, helping them climb the steps and in a certain way leveling them with their peers. This is where mentorship is the most significant to me when mentors help mentees navigate the difficulties of the academic world so that they can be their full selves and dedicate themselves to doing the best in their careers.

Salma: Postgraduate students typically come from different educational and socio-cultural backgrounds. As such, their journey through graduate school and integration into research life never takes a linear trajectory, but rather a combination of arcs, loops, and terrains. This makes it impossible to cover different learning needs through structured curricula alone. Mentorship is an excellent tool that can succeed where other tools fail. For instance, a mentorship relationship holds the potential to facilitate a meaningful exchange of ideas on life and career-related subjects that may prove challenging to broach within other types of academic relationships. In such cases, a mentor acts as a guiding force who provides valuable insights into navigating complex professional landscapes and offers constructive feedback and advice to help mentees overcome challenges and achieve their goals. The depth of these interactions and nuances of experiences shared often inspires growth and development for both parties. Furthermore, the benefits of mentorship extend beyond the academic realm; it also fosters growth in personal and professional spheres by opening doors to new networks, providing opportunities for career advancement, and enhancing mentees' self-confidence and esteem. The same outcomes might not be easily achieved through a typical supervisor-student relationship, where the inherent power dynamic and plausible conflict of interest might be a cause of distance, hesitance, or resistance, which could adversely impact such exchanges.

2. What have been some impactful mentorship strategies we have used or experienced?

Progna: As an early career researcher, I have found that one of the most impactful mentorship strategies is active listening. When I am mentoring a student or junior colleague, I make sure to listen carefully to their concerns, ideas, and goals. This allows me to provide guidance and support that is tailored to their individual needs, rather than simply giving generic advice.

Another important strategy is setting clear expectations and goals. I make sure that my mentees understand what is expected of them, and what they can expect from me as their mentor. This includes setting clear timelines, goals, and objectives, as well as regularly checking in to see how they are progressing.

In terms of my own experience as a mentee, I have been fortunate to have mentors who were supportive, encouraging, and willing to invest time and effort in my professional development. They provided me with opportunities to grow, challenged me to think creatively, and offered constructive feedback that helped me improve.

I believe that effective mentorship involves building strong relationships based on trust, respect, and mutual understanding. By listening carefully, setting clear goals, and providing guidance and support, mentors can help their mentees to reach their full potential and achieve their goals."

Elisabeth: Just like every relationship a mentorship is something highly individual. It needs to match the personalities, needs and abilities of the personalities involved. Nevertheless, I think there are some basic requirements such as trust, approachability and responsibility that will be part of every effective mentorship, and there are strategies which help to fulfill these:

- Having planned and spontaneous exchange between mentee and mentor. Planned meetings are important to structure the mentorship and to make sure you keep in contact, but often spontaneous meetings are much more helpful to solve problems when they are pressing or to talk about issues in a less formal atmosphere.
- Talk about your mentorship. Make clear what you expect from each other and what your respective responsibilities in your mentorship are. Find ways to communicate when you want something to change in your relationship.
- Talk about the mentee's plans for the future, how to get there and break it down into concrete tasks. Mentors are particularly important in transitioning phases of the mentee's career. The mentor can help to identify concrete and realistic goals as well as potential threats and shortcomings. It helps to write these points down and re-evaluate in your next meeting.
- Take inspiration from the corporate world. As scientists we usually stress that we do things differently than in the corporate world and that our ecosystem needs more freedom and less structure. While we often consider the more formalised leadership in companies as too hierarchical, some of the there established tools can be very helpful. For examples (bi)annual evaluation meetings can be useful to monitor progress and identify future focus points of the mentorship and it can be very helpful to use the SMART model when formulating goals.

Ernest

a. Strong, respectful and unbossed working and mentoring relationship
b. Active Listening and open to diverse perspectives
c. Collaboration and knowledge sharing
d. Nurturing environment & encouraging/fostering Diversity, Equity and Inclusion
e. Offering constructive feedback and challenging opportunities for growth & development beyond comfort zone

Helena: I, fortunately, had the opportunity to work in different research groups, which were very good working environments overall. Of course, no place is perfect, as there is always a place for improvement. As a young professional who wants to lead my own group in the future, I believe that a good strategy is to critically evaluate the working environment where you are and how people interact to try to absorb the positive points and try not to reproduce the negative ones.

While mentoring, it is important to be aware that every person is unique and that mentorship must be adaptable to their needs. Some students, for instance, may require more guidance and attention, while others are more independent. Some may have the best family support and structure to achieve their goals, while others may face unthinkable adversities that might impact their performance. A good mentor should be aware of these nuances to give guidance in a more personalized way.

Gabriele: In my opinion, one of the most effective mentorship strategies is leading by example. Throughout my career, I have always admired my mentors in many aspects, which pushed me to believe in them and in my own abilities. Their acceptance of failure, inspiring resilience, and willingness to listen to my problems and concerns have been precious lessons and memories for me. I feel grateful for the chance to learn by simply observing my mentors' attitudes. If I am who I am today, it is mainly because I followed in the footsteps of my mentors, who showed me the way.

As a mentor myself, I not only try to transfer what I have learned during my journey so far, but I also do my best to allow my mentees to express themselves: their personality, character, and attitude. From this perspective, a brainstorming session is a perfect environment where everything can be shared in a safe space. A brainstorming meeting always provides an opportunity to get to know people better, and understand their train of thought, and their mindset, which creates a human connection beyond a simple scientific plan. As a mentor, I always try to form a bond with those around me through empathy, as I have realized that it is more important to remember that we are human beings, regardless of any other label we may give ourselves.

Kevin: In my experience, some of the most impactful mentoring has come from informal discussions and observations rather than from formal mentoring events or meetings. For example, my PhD advisor challenged me to think about my career and the bigger picture of how science works in ways that were initially very unnatural to me. However, in hindsight, it was incredibly valuable. Additionally, reflecting on how different researchers do science or lead their groups has been helpful, especially when discussing these observations with mentors.

When it comes to making career decisions, I've found that having someone to bounce ideas off of can be extremely beneficial, even if the decision ultimately falls on me.

I've also found that mentoring others has helped me grow in many ways. Trying to transfer knowledge and help others has allowed me to reflect on my own experiences and gain a better understanding of where the mentee is coming from, including their goals and past experiences.

In my opinion, an important part of mentoring is understanding the mentee's perspective and goals. This helps tailor advice to the specific needs of the mentee and can lead to more effective mentorship. I did not find a silver bullet strategy that I can always use – I find mentoring to be highly individual and that I need to adjust my approach with every mentee.

Laura: A good mentor to me is someone who can listen and direct you to a solution or to the next step without pushing any particular agenda. They can be anyone but career-wise it has been helpful to me to have mentors one (or a few) steps ahead.

I am a mentee within a group established by the Women in Supramolecular Chemistry support group who connected myself and (at the time) 3 other PhD students with a mentor, we are all in the same field so there is a common ground but at different institutions such that we can discuss topics openly without judgement. This highly effective mentoring group began in 2019 and we have been meeting every month for 3 years (and continuing). Although the group has evolved, I still regard this group as my most effect professional support group. I think the key to our successful ongoing mentor-mentee relationship results from having clear expectations and boundaries, including how long and how frequent the meetings will be; we commit to 1 hour per month and make sure we never run over time, not only is this a manageable length of time to commit to but also ensures we stick to topic and have a productive meeting.

I believe it is also necessary to manage your own expectations of your mentors; know that they will not (and should not) solve all your problems for you, and that everyone has formed their opinions based upon their own experiences and biases etc. My most effective mentors (and what I try and do as a mentor to other) know when to listen and when to give advice appropriately. Small group mentoring can be particularly good because more diverse perspectives are bought to the discussion.

Finally, being proactive with any 'homework' that results from mentoring can lead to a much more productive mentor-mentee relationship in which no one feels like it was a waste of time, and then follow up with results (e.g. doing the job application, having a conversation etc.).

Maha: Impactful mentorship strategies I have experienced below:

1) Consistent one-on-one meetings: I have consistent one-on-one meetings with my graduate advisors, and these are one of the most effective things in my graduate school. My advisors have always given me the liberty to set my own research agenda. These meetings are extremely sacred for me as I use them to set small research milestones for myself. I also sometimes use them as brainstorming sessions with my advisors. There were a handful of meetings when we only discussed the problems, I was facing in data-analysis and how to best address those issues. These check-ins have also taught me how to prioritize and manage multiple research projects.

2) Effective communication: Any strong relationship starts with effective communication. Communication that is open, respectful, healthy, honest, and frequent. As a mentee, I have learned a lot from the communication styles of my advisors. Effective communication also helps resolve conflicts and creates a positive working environment for everyone involved.

3) Compassion, kindness, trust, and admiration: As a mentee, I can say with 100% conviction that my mentors have deep compassion and kindness towards me, and this has inculcated trust in me for them. I trust their research advice. They trust my research work. When there is trust, you listen to each other. I think these qualities are extremely important in an effective mentorship relationship.

   The last quality I will list is admiration. By admiration, I mean looking up to your advisor's work/research with regard. I look up to my advisors as my role models in science and this has shaped my personality as to how I envision running my own research group one day.

4) Effective feedback: Effective feedback is: direct, clear, and timely. For example, when my advisors provide me direct and clear feedback on my manuscripts and presentations, I know what they are looking for and how best to implement their feedback. Additionally, they set realistic timelines for providing feedback which are mostly two weeks for first review. This direct, clear, timely, and frequent feedback has been instrumental in my development as a first author on research projects.

5) End-of-year assessment: I think it is important to set milestones and then look back at these after a set period. My end-of-year assessments with my advisors have been very helpful in reflecting on my achievements, and areas of improvements the prior year and then using those to prioritize goals for the next year. "

Sai:

1) Setting clear expectations: An effective mentoring relationship begins with a strong foundation. A key element of building that foundation is having a clear on-boarding procedure for mentees. Similar to how a company would welcome a new employee, mentors should create a welcoming and professional environment for their mentees. This is especially important in academic settings, where new students and postdocs can feel overwhelmed and unsure of how to navigate their new roles and responsibilities.

   The on-boarding process should begin with an orientation session where the mentor provides an overview of the lab or research group's culture, values, and expectations. This session should cover the mentor's expectations for the mentee's work, including the expected timelines, communication protocols, and feedback mechanisms. The mentor should also provide information about the resources and support available to the mentee, such as lab equipment, software, and funding opportunities.

   Another important aspect of the on-boarding process is setting clear goals and objectives for the mentoring relationship. Mentors should work with their mentees to develop a shared understanding of what the mentee hopes to achieve during their time in the lab or research

group. This could include specific research projects, skill-building opportunities, or professional development goals.

2) Encourage independence: Effective mentoring involves more than simply providing guidance and advice to students and postdocs. It is a complex process that requires the mentor to be attuned to the needs and goals of their mentees, and to have a clear understanding of how to provide the support and guidance they need to achieve their objectives. One of the most important aspects of effective mentoring is the need to encourage responsibility in students and postdocs. By empowering them to take ownership of their work and to manage their time effectively, mentors can help to build trust and foster a sense of independence that is essential for success.

   Encouraging independence means giving students and postdocs the freedom to find their own pace and to work in a way that suits their individual learning styles and goals. This means not micromanaging their work or constantly checking in on them to see what they have accomplished. Instead, it involves setting clear expectations and goals, providing regular feedback and support, and trusting them to take responsibility for their own progress. This can be challenging for some mentors, especially those who are used to more hands-on approaches, but it is essential for building trust and fostering independence.

   One way to encourage responsibility is to involve students and postdocs in the development of research projects and goals. This means setting clear objectives, but also giving them the freedom to explore their own ideas and to pursue research questions that interest them. This can help to motivate them and to give them a sense of ownership over their work, which can be a powerful motivator for success. Additionally, providing regular feedback and support, such as offering guidance on research methodologies, reviewing drafts of papers, and providing constructive criticism, can help to build confidence and encourage independent thinking.

3) Finding the right fit: Lab work can be challenging and demanding, requiring dedication, long hours, and attention to detail. Mentors need to understand that not everyone is cut out for lab work, and it's not a failure to acknowledge that. Therefore, it's crucial to find a good fit between mentors and their students or postdocs.

   Mentors need to understand the strengths, weaknesses, and personalities of their students and postdocs. This information can be obtained through discussions, observation, and regular feedback. When mentors know their students' personalities, they can identify the right approach and communication style that works for each student. Some students may prefer more hand-holding and constant check-ins, while others prefer autonomy and independence.

   Mentors should encourage students to rotate in different labs to find a good fit. This approach can help students identify the kind of research they are interested in and the type of work environment they thrive in. Mentors should also allow their students to pursue their interests and passions. When students are working on projects that they are passionate about, they are likely to be more productive and engaged.

Finding a good fit between mentors and their students or postdocs is beneficial for everyone involved. When the fit is right, students are more likely to be engaged and productive, and mentors are more likely to enjoy working with their students. It also helps to build trust and respect between mentors and their students or postdocs.

4) Embracing diverse perspectives: In recent years, there has been a growing recognition of the importance of diversity and inclusion in academic and professional environments. American universities have made significant efforts to increase the enrollment of students from underrepresented groups and underprivileged backgrounds. The result is that the students in the labs are starting to reflect that trend. This shift in the student body's demographic makeup requires a corresponding shift in mentorship practices to ensure that everyone feels accepted and welcome.

   A good mentor should be aware of the challenges that students from diverse backgrounds may face and should take proactive steps to address them. Mentors should strive to create an inclusive environment that values diversity and promotes equity. This means creating an environment where students from different backgrounds can thrive and where their unique perspectives and experiences are valued. By doing so, the mentor can ensure that all students have equal access to opportunities and resources.

   Additionally, it is essential for mentors to be aware of the financial challenges that many students face. Graduate students, in particular, often need a fellowship to afford their programs. A good mentor should do everything they can to help their mentee secure funding. This could mean connecting them with funding opportunities, providing them with advice on how to write a grant proposal, or introducing them to potential funders. By doing so, the mentor can help ensure that their mentee can focus on their research and academic pursuits without the added stress of financial concerns.

   Another way that mentors can embrace diversity is by encouraging their mentees to pursue interdisciplinary research. This involves bringing together individuals from different fields of study to collaborate on research projects. Such collaborations can bring together a range of perspectives, allowing for a more comprehensive understanding of complex issues. Moreover, interdisciplinary research has been shown to lead to more innovative solutions to problems. By encouraging their mentees to pursue interdisciplinary research, mentors can help promote diversity in academic research and foster a more inclusive and collaborative environment.

5) Forging a personal connection: Mentors should make a point to meet with their mentees on an individual basis regularly. These meetings provide an opportunity for mentors to build a rapport with their mentees, discuss their academic progress, and offer guidance and support. Moreover, mentors should also be open to discussing personal issues with their mentees. While the primary focus of these meetings may be on academic and career-related matters, personal conversations can be valuable for the mentee's overall well-being.

   It is common for university students to experience depression, anxiety, and other mental health issues. The pressures of academic work, combined with the challenges of navigating

the transition to adulthood, can be overwhelming for some students. As such, it is crucial for mentors to be aware of the mental health challenges that students may face and be available to provide support.

In these one-on-one meetings, mentors should make it clear that they are available to listen to their mentees' personal concerns. By creating a safe and supportive environment, mentors can encourage students to open up about their problems. Once the students feel comfortable enough to share their personal issues, mentors can provide guidance and advice or refer them to appropriate resources such as counseling services.

Additionally, mentors can use these conversations to offer advice on work-life balance. It is essential for students to prioritize their mental health and well-being alongside their academic and professional pursuits. Mentors can offer suggestions on how students can manage their workload, develop good study habits, and practice self-care.

6) Embracing failure: As humans, we are all prone to making mistakes, and it is essential for mentors to acknowledge and share their failures with their mentees. Many times, students feel that they are the only ones struggling and that everyone else is achieving success effortlessly. By sharing their own failures and setbacks, mentors can create a supportive environment that encourages students to persevere despite setbacks.

   Sharing failures can also help students understand that failure is a natural part of the learning process. When mentors share their failures, they can explain how they learned from their mistakes and how they overcame the challenges they faced. This can help students develop a growth mindset and see failure as an opportunity for learning and growth rather than a setback.

   Furthermore, sharing failures can help mentors build trust with their mentees. When mentors share their own failures, they demonstrate vulnerability and authenticity. This can make students feel more comfortable opening up about their own struggles and concerns.

   It is important to note that sharing failures should be done in a constructive way. Mentors should not dwell on their mistakes or use them as an opportunity to vent. Instead, they should focus on the lessons they learned and how they overcame the challenges they faced. Additionally, mentors should be mindful of their mentees' emotional well-being and not share failures that may cause undue stress or anxiety.

Silvina: I have been on both sides of the spectrum--I have been a mentor (I taught 6th and 7th grade middle school science) and a mentee (during my graduate career). Some impactful mentorship strategies that I experienced as a mentee have been:

- Provided resources.
- Weekly 1 on 1 meetings.
- Feedback on both technical and soft skills.
- Room for creativity and discussion.

The aforementioned strategies are fundamental yet rather commonplace. One strategy that I know is not commonplace is liberty. I feel fortunate that my mentor was open to change the dynamic of the mentor-mentee relationship as I became a more "experienced" graduate student. As I became more independent, I started to set expectations for myself. It was a sense of freedom and creativity. Then, I would share my own expectation results with my mentor during our weekly 1:1 meetings. During my last four years of my graduate career, I decided to create a list of "Accomplishments in 20XX" and "Continue improving in 20X+1". For four consecutive Decembers, my mentor wrote a list of things "well done" that year and another list of things to "improve" the following year. This end-of-the-year assessment was instrumental to my growth as a young professional.

As a mentor, I have implemented the strategies my mentor used (above) for me. In addition to the ones I implemented on my own:

- Active listening.
- Set expectations.
- Guidance on professional and personal life.
- Constant communication, all plausible flatform.

To have a successful mentor-mentee relationship, it is important to be dedicated to your mentee ('active listening'). One needs to be truthful to oneself and genuinely provide positive feedback and constructive criticism, both in the professional and personal realm ('guidance on professional and personal life'). Thus, I believe it is important to be introspective and spend some time focusing on what one desires to achieve with the goal of having a mentee. Again, one needs to serve a purpose in a healthy, symbiotic relationship ('set expectations'). Frequency is key, so both the mentor and mentee have opportunities to follow-up and refresh expectations and materials discussed ('constant communication').

Lastly, to achieve effective mentorship, we need to have a long-term purpose, with the goal to grow endlessly. Quoting Barrette-Ng et. al, work, "commitment to mentorship is a commitment to lifelong learning within an academic micro-community." [Barrette-Ng, N., Nowell, L., Anderson, S. J., Arcellana-Panlilio, M., Brown, B., Chalhoub, S., ... & Wilcox, G. (2019). The mentorship guide for teaching and learning. Taylor Institute for Teaching and Learning Guide Series.]"

Olivia: Something I learned early on in my PhD program is that sharing moments of vulnerability as a mentor can have long-lasting effects on the outlooks of your mentees. When I began learning new data analysis methods as a first-year graduate student, I felt overwhelmed. As I watched Brandon—a senior PhD candidate who became my mentor—effortlessly scan lines of code and explain what they did, I couldn't help but think that I would never, could never reach that point of expertise. If brains are sponges, mine was oversaturated with new knowledge and unable to retain anything new. One day, while Brandon and I stared at my computer screen, my carefully guarded anxiety that I wasn't cut out for this type of research snuck out. I don't remember exactly what led up to this moment, but I clearly remember that Brandon looked at me knowingly and said, "You'll get there. First year is like drinking out of a firehose." He was months away from defending his PhD, but five years earlier, he had felt just like I did? That seemingly simple statement was empowering. I wasn't falling behind; I was right where I needed to be in my PhD journey.

Over the next several years, I had somehow accrued enough knowledge to become the go-to person for early-stage graduate students who had questions about the data analysis software. Whenever I

sat down with a student to look at their data, I made an effort to tell them what I had been told, that learning to use the software was like drinking from a firehose. I remember one student's facial expression immediately relaxed when I said that, much like I'm sure mine did three years before.

After finishing my PhD, I switched from studying chemistry via telescope observations to working in a lab for the first time in more than five years. I felt much like I did during my first year of grad school. I wasn't even going to apply for the fellowship to work in a lab because I clearly was not qualified to do experiments, but some nudging from my PhD advisor convinced me to give it a try anyway. In the first week or so of my appointment, my postdoctoral advisor (who had made a similar switch of fields) told me how she understood such a big change could be overwhelming and that I could reach out to her to talk through it if I needed. That type of vulnerability was again empowering, especially coming from someone in a supervisor role. In the past, scared to show that I knew less than everything, I had kept lists of questions about concepts I didn't understand to try to find answers on my own time. But knowing the expectation was that I started out knowing very little about this new project, I felt confident enough to ask questions in real time, and my learning has been much faster because of it.

In research, we place so much emphasis on people being "experts" in something that I think we sometimes forget that we can't be experts in all the things. Reminding mentees that they aren't expected to know everything (or perhaps even anything) when they start out can help them feel more secure in their pursuit of knowledge, empowering them to spend their energy on stumbling through research rather than on pretending they are going through the process flawlessly. Instead of perceiving asking questions as a sign of weakness, mentees can see asking questions as an opportunity for growth.

Stanna: Some of the best mentors I have had are excellent listeners. Whether I was describing a time I struggled, or an action that I believe they could have done differently, I have had mentors not only listen, but value what I was sharing.

The "hidden curriculum" of academia can be hard to navigate without a mentor. Throughout my journey, I have had mentors share resources and knowledge with me; as I advance, I compile and share information as well. I continually make myself aware of opportunities and resources, knowing that although I may not have all the answers as a mentor, I can hopefully point a trainee in a direction that will help them.

Julio: Impactful mentorship for me happens when the mentor-mentee connection is somehow personal, meaning that they acknowledge each other not only as professionals but also as individuals. The mentee needs to feel comfortable expressing their background and personality, and this is when they are their true selves and exceed in what they do. Of course, this is easier when the mentor is comfortable and open about themselves. As an LGBTQ+ scientist, I was lucky to find mentors who were allies, with whom I could always be myself and express my identity and personality. We know that LGBTQ+ scientists are less likely to be retained in STEM than their non-LGBTQ+ colleagues (DOI 10.1126/sciadv.aao6373) and that LGBTQ+ scientists are more productive and successful when they can express their sexual and gender identity in the work environment (DOI 10.1042/BIO20200024). I am certain that the support and confidence that I have always had from my mentors played and still play an essential role in my retention and

achievements in STEM and academia. As a mentor, I aspire to build a relationship of trust and respect with my mentees and acknowledge their personality, diversity, and uniqueness, as this is from my experience the foundation for a truly impactful mentorship experience.

Salma: Effective mentorship relationships require trust, mutual respect, and clear communication between mentors and mentees. They also require certain established protocols, a clear definition of boundaries and an understanding of roles, expectations, responsibilities and goals. These elements contribute to the foundation of successful mentorship relationships that are focused on development and growth, while also creating a safe space for mentorship.

I have been fortunate enough to engage in a diverse range of mentorship styles at different stages of my personal and professional career, with each relationship bringing forth unique experiences and perspectives that responded to different needs and played a vital role in shaping who I am today. Having a female mentor allowed me to be more vulnerable and realistic as I plan my career aspirations. Male mentors on the other hand helped me build my self-confidence while focusing on being more productive and professional. Alternatively, in-person interactions facilitated the development of long-lasting and robust relationships with my mentors and junior mentees, whereas virtual mentoring provided opportunities to connect with accomplished individuals possessing innovative viewpoints that may have otherwise been challenging to access.

My experiences taught me to appreciate the mentee-initiated mentorship, where I took initiative in seeking guidance and support from seasoned professionals, as well as the mentor-initiated relationships in which seniors and more experienced mentors took me under their wings to nurture my skills and help me advance in my career. I have mostly appreciated structural mentorship, which was facilitated through formal programs and involved pre-defined roles, programmatic goals, and set timelines. I found that the mentee-initiated style often suits more extroverted individuals who are confident in initiating conversations and building relationships. Analogously, effective mentor-initiated relationships require mentees to be more receptive and open-minded, with a willingness to learn from their mentors and appreciate their guidance, commitment and time. Comparably, the structural style can suit either personality. It might however be more restricted by time and requires mentors and mentees to regularly review and assess the progress made in their relationship and make adjustments where necessary.

3. How can we create a chemistry community that values and prioritizes effective mentorship?

Progna: Creating a chemistry community that values and prioritizes effective mentorship is crucial for the success of the field and the growth of the next generation of scientists. One way to achieve this is by recognizing and celebrating exceptional mentorship. Institutions can incentivize and reward mentors who demonstrate effective mentoring by providing grants, fellowships, or awards. Moreover, it is important to provide mentorship training to both new and experienced faculty members to ensure they are well-equipped to be effective mentors.

Another key strategy to create a culture of effective mentorship is by promoting an environment that values collaboration and communication. Chemistry departments can organize seminars, workshops, and events that encourage meaningful interactions between faculty, students, and

postdocs, and offer opportunities for mentor-mentee relationships to develop. Establishing a network of mentors and mentees within the department can also be helpful in providing a supportive community and encouraging mentorship from various perspectives.
Lastly, it is important to provide students and postdocs with the resources and support they need to excel. Creating a supportive environment can be achieved by offering workshops on professional development, financial planning, and work-life balance, as well as providing access to counseling services and support groups. By prioritizing effective mentorship and creating a supportive community, we can ensure that the chemistry field continues to grow and thrive.

Elisabeth: Being a good mentor is a skill that needs to be developed. As many aspects of leadership, until now it is undervalued in our education and curricula. That not only makes it less likely for people to recognize that effort needs to be invested in order become a good mentor, but also makes it less realistic that they find time and energy to actually do that. With mentoring programs becoming more and more common, it might be time to also implement programs in which potential mentors get time, support and resources to develop their skills.

In the scientific community we still have the ideal of the genius scientist who can do it all on their own and "just like that" develops the perfect solution to whatever problem they face. Sometimes highly successful scientists talk about how much they enjoy being a mentor, but only very rarely it occurs that they talk about their experiences as mentees, about the times when they struggled and needed some support to progress. We all can contribute to a change of culture by doing that ourselves, talking openly about the problems we had and how we could benefit from the experience and support of others. This will give evidence of the importance of mentorship and help younger scientists to understand how many puzzle pieces and contributions are needed for a good scientist to develop. Equally important it will reduce stress and imposter syndrome by lowering the imaginary bar to a more realistic height.

Ernest: I believe creating a chemistry community that values and prioritizes effective mentorship starts with awareness on the importance and benefits of effective mentorship to both parties involved. As I've previously mentioned, I believe mentorship is a two-street, so I think educating and informing the chemistry community about the importance of mentorship not only to the mentee but also to the mentor is a good way to begin creating a community that values its impact. Another thing to consider is providing mentorship training courses and resources to all parties involved to prepare and educate everyone about the goals and expectations of effective mentorship for a rewarding and positive experience. Academic conferences, symposia and networking opportunities should have a component on effective mentorship, collaboration and engagement as part of their program. Lastly, fostering a nurturing, diverse, and unbossed environment where everyone is valued and feels comfortable to share their perspective is crucial in creating a sustainable community for effective mentorship.

Helena: The first thing to do is discuss mentorship itself. Mentorship strategies are not typical subjects that are brought up and debated or trained on how to do it best. Fortunately, these points are being increasingly discussed, and mentorship training is being offered and encouraged in several universities. Even more important than providing training to the mentors is allowing mentees to share their feedback safely, without fear of retaliation. Mentor-mentee must be,

indeed, a two-way relationship, and both parties should be open to this exchange for it to succeed.

People usually associate mentors with people very close to us or our work/project, but sometimes having input from someone outside this circle might be beneficial to give and get unbiased advice. Therefore, beyond someone's personal network, virtual platforms have been developed and encouraged by some societies and communities to facilitate such exchanges between different mentors and mentees. Getting involved in these new ways of mentoring can be enlightening for both parties.

Gabriele: I believe we can create a chemistry community that prioritizes effective mentorship by promoting constant communication between mentees and mentors. In particular, evaluating mentees' performance is essential to help them improve their skills and attitudes. Advertising courses on this topic would be a valuable way to create connections and improve working relationships. In fact, giving mentors and mentees the right tools to communicate effectively by sharing opinions and providing feedback would dramatically increase productivity and research quality in both academic and industrial settings. Other opportunities, such as mental health counseling or psychological support, without stigmatizing or downplaying individuals' discomfort, would open up a new way of seeing researchers and workers, propelling a cultural revolution, and making effective mentorship a natural consequence of this shift.

Moreover, I believe it is important to inform and train people to use social media professionally. Gaining expertise on how to approach social media would not only be important for scientific communication but also to acknowledge and express different opinions in the STEM community.

To conclude, we can create a more accountable chemistry community centered on mentorship by building an empathic connection across different generations of scientists, which will lead to a more united and reliable scientific society.

Kevin: In my experience, formal programs that assign mentors to Ph.D. students can be a good starting point, but it is not a complete solution to effective mentorship.

One impactful strategy is to emphasize how being a mentor can contribute to personal growth. Mentoring provides an opportunity to reflect on one's own experiences and can lead to new insights and perspectives.

Another strategy is to start mentoring early, even in undergraduate programs. For example, having students mentor younger students can not only benefit the mentee, but also help the mentor develop valuable skills.

Finally, it may be beneficial to reward mentorship and make it part of the tenure evaluation process. This can help incentivize effective mentorship and encourage more senior researchers to invest in mentoring the next generation."
Laura: I believe systemic change in opinion will have to happen and then passed down to the next generation of leaders and mentors in order create a chemistry community that values and prioritizes effective mentorship. Despite some PIs having a bad reputation when it comes to

mentoring, students and postdocs may still want to work for them to benefit from the prestige that may come with a particular group. Not all believe that effective mentorship is critical for the success of research and therefore don't prioritise it. Unfortunately, this may result some people having a bad PhD experience, or even dropping out, and seeking mentorship outside of your institution may not be possible.

One strategy may be to prioritise collaboration over competition, at every level of academia; from members within the same group to research groups at different institutions. I believe this has to start with each individual research group fostering open communication and collaboration. Sharing knowledge, teamwork and group problem solving should be encouraged within a forum that allows everybody the opportunity to have a voice. This fosters a professional working environment in which members feel supported. I believe when the benefits of effective mentorship acknowledged, such as increased productivity, the leaders within the academic chemistry community will be encouraged to carry this attitude forward.

Maha: There are very few URMs in my field of research and a way to address this issue is their effective mentorship. I think it is extremely important to draw awareness to this issue in academic circles. An idea is to have a session or two on this topic at the American Chemical Society Fall Meetings. It will be great to see chemistry researchers around the world listen to each other's mentoring experiences and develop strategies for their mentor-mentee relationships.

Silvina: A chemistry community that prioritizes mentorship must show utmost importance towards mentor-mentee relationships. Although counterintuitive, "making" science should be placed in the "backburner", prioritizing those first- and second-year graduate students.
In addition,a well-developed and strategized system must be already in place, implemented by curricula. Also, I believe that an effective mentorship is one that prioritizes communication; thus, frequency in meetings is primordial. One-on-one meetings are key to discuss both technical and professional development progress. "

Olivia: When I was working on my PhD, students at my institution frequently requested that the department provide mentorship training to faculty. The idea was not welcome, at least at the time, for several reasons: it was expensive, it was unnecessary, it was even insulting (because faculty wouldn't be hired if they weren't qualified to be mentors). Instead, how faculty learned about mentorship seemed to be limited to pulling from one's personal experiences of what worked and what didn't work, which is strongly dependent on the person. (There were, however, a few people who sought out mentorship training on their own!) In response, students suggested that committee meetings include portions during which the mentor was required to leave the room to mirror portions during which mentees were required to leave during discussion. In these aspects absent of the mentor, students would be able to raise concerns about mentorship practices and get feedback on how to address said concerns. Again, the idea did not gain traction.

A chemistry community that values and prioritizes effective mentorship needs to have opportunities in place for mentors or future mentors to engage in mentorship training and receive feedback. Effective mentorship, like any other facet of research or teaching, is a skill that requires learning and practice. These are not easy things to implement, even with departmental support, because they come with added time requirements for faculty and (perceived and real)

risks for graduate students. Effective training requires dedicated time beyond a single seminar, for example. Graduate students and faculty peers who would provide feedback on mentorship practices, especially in small research groups or departments, may fear retaliation and decline to provide feedback on areas that would benefit from changes. To prioritize effective mentorship, we need to shift to a growth mindset in which developing an effective mentorship is seen to benefit everyone (e.g., by giving graduates better outcomes, increasing research productivity, having a more collaborative and collegial environment versus a disproportionately skewed power dynamic) rather than as a criticism of how people lead.

Stanna:

1) Creating/providing access to formal training opportunities: It is intriguing that although much of a PI role involves management (of trainees, research projects, funding, etc.), this skillset is often not a significant criteria for the hiring process. Furthermore, formal avenues for obtaining management training in a scientific setting are not as widely known or accessible. Much of a PI's management training is done "on the job," and herein lies the dissonance–the criteria valued at the time of hiring does not fully encompass what the role actually requires, and trainees pay the price for this inexperience. Institutions need to value mentorship when hiring, but also support mentors in obtaining training for these skills.

   Additionally, graduate students and postdocs are usually expected to train other graduate students or undergraduates without any guidance on how to do so; having formal training or best practices for this type of mentorship could mitigate harm done by ineffective mentoring.

2) Increased accountability: Academic institutions often claim that they prioritize effective mentorship, but all too often harbor and promote mentors/PIs that display recurring patterns of harm/abuse; the power imbalance is skewed such that trainees are often reluctant or afraid to give feedback, and even when they do, institutions are slow to act on it, if at all. If we want to create a community that values and prioritizes effective mentorship, we need to demonstrate that it is something to be taken seriously and acted upon, and not merely hollow statements without any accountability.

Julio: Although most of us become mentors at some point in our careers as chemists, we are not trained for it and are many times not aware of good practices. I believe the chemistry community needs to acknowledge mentoring as an essential skill, especially in academia (where we all spend many years of our lives and when we need mentors the most). We need to share good and bad experiences and highlight their impact on the mentees' careers. I believe quantitative research that draws relationships between mentorship experiences and mentees' career paths could be an effective way to highlight this impact and get the community to better value and prioritize effective mentorship.

Salma: By promoting and encouraging mentorship, STEM communities can foster an environment that supports the growth and success of all members. This can ultimately lead to increased diversity and inclusivity within the STEM community, as well as better-prepared professionals who are equipped with valuable skills and knowledge to contribute to the

advancement of scientific knowledge and technological innovation. Engaging in academic discussions regarding the advantages of fostering functional mentorship alliances, particularly for marginalized individuals within the STEM field, constitutes an initial measure. Senior STEM professionals should be encouraged to lead by example and serve as role models for mentorship in the community. Mentors should also be encouraged to communicate the advantages they acquire by mentoring others. Ultimately, this would help counteract the prevalent notion that only mentees benefit from mentorship relationships. Such action can also stimulate increased involvement of other mentors in similar activities.

Analogously, incorporating productive mentorship connections into STEM education as an essential component, from the high school level onward, could potentially foster significant progress within the community. This would also help students and EMCRs gain a clear understanding of the distinctions among supervisory, mentorship, and training relationships and their respective expectations.

Encouraging and facilitating mentorship programs, along with establishing accolades that incentivize mentoring endeavors among diverse professional groups and organizational departments, should be given significant attention as well. Moreover, creating opportunities or platforms to evaluate the effectiveness of mentorship programs and providing feedback to mentors could fortify the implementation of functional mentoring alliances.

4. Anything else you would like to share?

Progna: As an early career researcher, I have grown both personally and professionally while cultivating mentorship and diversity ideals and practices. My growth as a scholar and mentor has been closely tied to my values of inclusivity, diversity, and support, and I am constantly striving to create a lab community that reflects these ideals. Initially, I had a traditional approach to mentorship, where I would provide guidance and support to students and postdocs in the lab. However, I realized that my role as a mentor goes beyond technical guidance and extends to the personal and professional development of my trainees. I actively seek to create a supportive and inclusive lab environment where everyone feels valued, heard, and supported. This involves being available to my trainees, providing feedback on their work, and encouraging them to pursue their interests and goals.

Additionally, I have recognized the importance of diversity and inclusion in the lab and have implemented practices that prioritize equity, such as creating opportunities for underrepresented students and fostering a culture of respect and inclusivity. I have also worked on developing a mentorship network that extends beyond the lab, connecting trainees with mentors and resources that align with their interests and career goals.

Laura: Resources (That I am aware of/have used – sometimes you can't find a mentor but other resources can be just as useful throughout a PhD)

Mentoring:

- The Royal Society of Chemistry: https://www.rsc.org/careers/cpd/mentoring/

- Women in Supramolecular Chemistry: Mentoring | Women in Supramolecular Chemistry (womeninsuprachem.com)
- The Thesis Whisperer: The Thesis Whisperer

Books:

- "You Are Not a Fraud: A Scientist's Guide to the Imposter Phenomenon" By: Dr Marc Reid. You Are Not a Fraud: A Scientist's Guide to the Imposter Phenomenon — Marc Reid, PhD (dr-marc-reid.com)
- "Managing your mental health during your PhD" By Dr Zoe Ayres. Home | Mental Health in Academia (zjayres.com)

Websites:

PhD planning & Viva prep:

- Thinkwell: Tools to give to your student (ithinkwell.com.au) & PhD Toolkit (ithinkwell.com.au) & PhDVivaQuestions.pdf
- Viva Survivors: Viva Survivors – Daily viva help for PhDs (viva-survivors.com)

Planning Reactions:

- Reaxys: https://www.reaxys.com/#/login
- CAS SciFinder: https://scifinder-n.cas.org
- Master Organic Chemistry: https://www.masterorganicchemistry.com
- LibreTexts Chemistry: https://chem.libretexts.org/Ancillary_Materials/Demos_Techniques_and_Experiments/General_Lab_Techniques
- Not Voodoo X.4: http://www.chem.rochester.edu/notvoodoo/
- The Schlenk Line Survival Guide: https://schlenklinesurvivalguide.com/
- Chemistry Views (Chemistry European Journal): Tips and Tricks for the Lab: Column Choices – ChemistryViews & Tips and Tricks for the Lab: Column Troubleshooting and Alternatives - ChemistryViews
- Data Organic Chemistry: NMR Spectroscopy :: 1H NMR Chemical Shifts (organicchemistrydata.org)
- Chemtips: Chemtips | Because Organic Chemistry is Hard Enough (wordpress.com)

Dealing with Literature:

- Feedly: https://feedly.com/i/my
- Genie: genei | AI-powered summarisation & research tool
- Endnote: EndNote
- Zotero: Zotero | Your personal research assistant
- Mendeley: Mendeley - Reference Management Software

Other:

- ACS Reviewer Lab: https://institute.acs.org/courses/acs-reviewer-lab.html
- ACS webinars: https://www.acs.org/content/acs/en/acs-webinars.html
- My Green Lab: https://www.mygreenlab.org/ambassador-program.html
- Scientific Writing: I focus and write | Scientific Writing Online Courses

Silvina: To my scientific community, whether you are an undergraduate student, starting your career or a tenure professor, close to retirement - is to become a mentor. Mentoring relationships are extremely fulfilling, where both the mentor and mentee learn from each other. My advice is to nature and appreciate this unique symbiotic relationship.

Author biographies

Jacky Deng (University of Ottawa)



Jacky Deng (he/him) is a PhD Candidate working with Prof. Alison Flynn at the University of Ottawa (uOttawa). Jacky's research focuses on the experiences of learners who speak English-as-an-additional language in chemistry, with implications for how individuals and institutions can better support learners from diverse backgrounds in science. As a graduate student, Jacky has published peer-reviewed articles and book chapters in chemistry and science education (*e.g.*, *Journal of Chemical Education*, *International Journal of Science Education*), and has presented at various local, national, and international conferences and seminars. Jacky was selected to be a 2022 CAS Future Leader and was the inaugural J.C. Tito Scaiano Graduate Student of the Year at uOttawa in 2021. He was also a recipient of a SSHRC Vanier Scholarship in 2021—Canada's most prestigious graduate scholarship. Outside of his research, Jacky has been engaged in various national initiatives and committees to improve equity, diversity, and inclusion (EDI) in Canadian chemistry and higher education, including work conducted with the Canadian Society of Chemistry's Working for Inclusivity, Diversity, and Equity (WIDE) Committee and the Society of Teaching and Learning in Higher Education's (STLHE) EDI Task Force. He loves teaching and was selected as the 2020 Teaching Assistant of the Year by uOttawa's Science Students' Association.

Progna Banerjee (Argonne National Laboratory)



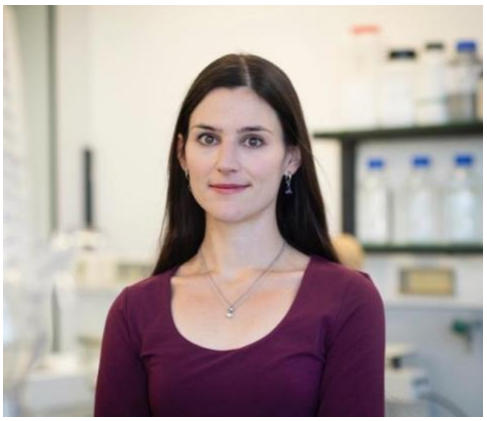
Progna Banerjee is a postdoctoral appointee at Argonne National Laboratory, where she works on interdisciplinary research in experimental condensed matter physics, materials chemistry, and electronics. She received her Ph.D. in Physics with Materials Chemistry and M.S. in Strongly Correlated Condensed Matter from the University of Illinois at Urbana-Champaign in 2018 and 2014, respectively, after earning her M.Tech. in Solid State Technology and M.Sc. in Physics from the Indian Institute of Technology, Kharagpur.

During her Ph.D., Progna used a template-based synthetic technique called "cation exchange" in chalcogenide nanocrystals from quantum dots to establish their unusual structural arrangement, which caused exhibition of room-temperature superionicity and topological insulating properties. At the University of Texas, Austin, where she was an NSF MRSEC postdoc fellow, she investigated light-matter interactions in bioinspired nanostructures. She has led materials sub-teams and worked on multi-university research projects with a focus on understanding the mechanism of light interaction in leafhopper-insect derived nanostructures called brochosomes with inherently high antireflective properties ascribed to their buckyball-like structures. While at the Lawrence Berkeley National Laboratory, where she was an EBI-Shell ALS affiliated postdoc fellow, she worked on battery science.

Currently, Progna is part of an interdisciplinary team at Argonne National Laboratory looking to discover nanomaterials using Artificial Intelligence/Machine Learning assisted colloidal synthesis and post-synthetic cation exchange combinatorics using an autonomous liquid-handling platform. Her research focuses on the expansion of the nanomaterials library into unconventional metastable phases and studying the emergence of various physicochemical properties in these systems.

Progna's interdisciplinary research background, particularly in the areas of experimental condensed matter physics, materials chemistry, and electronics, gives her a unique perspective in her field. She has authored several publications in Nature and ACS family journals, and has received recognition for her work, including awards for her research dissemination & academic leadership efforts, and presentations at international conferences.

Elisabeth Kreidt (Technical University Dortmund)



Already during her undergrad studies Elisabeth got fascinated by the unique physical properties of the lanthanoids and she then decided to study them in depth during her PhD. In the group of Michael Seitz she developed new functionalised lanthanoid cryptates, a class of structurally particularly well defined and stable lanthanoid complexes. Apart from using these special complexes for the realisation of a molecular nanocode she studied their photophysical and magnetic properties, including their circularly polarised luminescence. To broaden her knowledge in organic chemistry and to learn more about the toolbox about supramolecular chemistry she then joined the group of Dave Leigh in Manchester for her postdoc. There she realised and studied molecular knots of the size of small peptides and worked on the realisation of a chemically fueled unidirectional molecular motor. She then secured the scholarship of the state of North Rhine-Westphalia for scientists returning to Germany and since early 2022 she is setting up her independent research group. In her work she combines her knowledge about lanthanoid coordination chemistry, the physical properties of lanthanoids and supramolecular chemistry to develop systems which allow to dynamically control and manipulate the photophysical and magnetic properties of the lanthanoids. Aim of this work, among others, is to develop new tools for (bio)medical research, but also to improve the fundamental understanding of these properties. Elisabeth was elected a 2022 CAS future leader and is member of the young college of the North-Rhine-Westphalian academy of sciences and arts.

Ernest Awoonor-Williams (Novartis Institutes for Biomedical Research)



Ernest completed his B.Sc (Hons) degree in Chemical Physics at Mount Allison University under the supervision of Prof. Khashayar Ghandi in 2014. Ernest's honors research focussed on muonium free radical chemistry in green solvents, particularly supercritical carbon dioxide. This research combined both computational and experimental approaches, with the latter carried out with researchers at the TRIUMF National Laboratory in Vancouver, BC. In September 2014, Ernest moved to St. John's, NL to pursue graduate studies in Chemistry at Memorial University of Newfoundland. He completed his Ph.D. in Computational Biochemistry in May 2020 under the supervision of Prof. Christopher Rowley. His PhD research explored the use of computational methods in modeling the covalent modification of druggable cysteines in enzyme targets. As a Ph.D. student, Ernest was a visiting research scholar at the University of Ottawa studying enzyme inhibition by targeted covalent inhibitors under the supervision of Prof. Jeffrey Keillor. During his Ph.D. Ernest was awarded the prestigious NSERC Vanier Canada Graduate Scholarship, IUPAC Young Chemist award and an Advanced Research Computing Fellowship. In 2021, Ernest joined the Novartis Institutes for BioMedical Research (NIBR) in Cambridge, MA, as an Innovation Postdoctoral Fellow. At NIBR, he is exploring in silico approaches towards drug discovery, such as the impact of mutations in drug binding and pathogenesis.

Helena Mannochio Russo (University of California San Diego)



Helena Mannochio Russo received her B.Sc. degree in Chemistry from the University of Campinas (2015) and her M.Sc. and Ph.D. in Chemistry from São Paulo State University (2018,

2022). She carried out two internships at the University of Geneva, the first during her undergraduate studies and the second during her master's. During her Ph.D., she was a visiting scholar at Skaggs School of Pharmacy and Pharmaceutical Sciences (University of California San Diego), funded by the Fulbright Commission. During her undergraduate and master studies, she specialized in the phytochemistry of Brazilian plants. During her Ph.D., she performed mass spectrometry-based metabolomics and chemosystematics studies of Brazilian plants and investigated natural products with larvicidal activity against vectors of tropical diseases. Currently, she is a postdoctoral researcher at the University of California San Diego in Prof. Dorrestein's group, where she is interested in mass spectrometry repository-scale analyses, data science, and the chemistry of natural products. Helena was selected to attend the 71st Lindau Nobel Laureate Meeting (dedicated to Chemistry, 2022) as well as elected a 2022 CAS Future Leaders by the American Chemical Society.

Gabriele Laudadio (Scripss Research Institute)



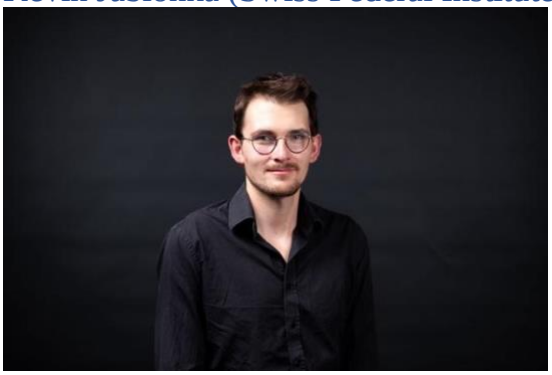
Gabriele Laudadio was born on July 12th 1991 near Teramo (Italy). He received his master degree in Organic Chemistry from the University of Pisa in 2016, under the supervision of Prof. Adriano Carpita. From May 2016, he started a Ph.D. project in the Micro Flow Chemistry and Synthetic Methodology at the Eindhoven University of Technology (Netherlands), under the supervision of Prof. Timothy Noël. The primary focus of his research was the application of Green Chemistry to improve Organic Chemistry methodologies, exploring the possibility to obtain molecules of interest faster and in a more efficient way, combining Continuous-Flow microreactor technology with electrochemistry and photochemistry. After a brief period as postdoctoral fellow and group leader in Prof. Noël's group at the Van't Hoff Institute for Molecular Sciences in the University of Amsterdam, he joined Prof. Baran's group at The Scripps Research Institute as a Hewitt Foundation Fellow. At Scripps, he is exploring the electrochemical functionalization of heteroarenes, to generate C(sp3)-C(sp2) bonds in a faster and efficient way. His contribution on flow chemistry and electrochemistry was recognized with several awards, including the CAS Future Leaders (2022), the IUPAC-SOLVAY International Award For Young Chemists (2021) and the IUPAC-Zhejiang NHU International Award for Advancements in Green Chemistry (2021).

Júlio Terra (Swiss Federal Institute of Technology)



Júlio was born and raised in a village in the Brazilian countryside, he obtained his B.Sc. degree in Chemistry in 2015 from the Federal University of Lavras (Brazil) with an undergraduate exchange period at the University of Toronto (Canada) in 2012-13, when he worked at the laboratory of Professor Bob Morris. In 2016, Júlio obtained his M.Sc. degree from the Federal University of Minas Gerais (Brazil) under the supervision of Professor Flávia Moura. He then joined the group of Professor Audrey Moores at McGill University (Canada) to pursue his Ph.D. degree in Green Chemistry and Catalysis, which was obtained in 2023. During his Ph.D., Júlio worked on the synthesis and nanodesign of silica-based structures that were used to make catalytic processes greener and more sustainable by exploring photocatalytic properties and nanoconfinement effects. Júlio is now a Banting post-doctoral fellow at the Swiss Federal Institute of Technology in Lausanne (Switzerland), where he designs nanocatalysts using atomic layer deposition to be used in biomass valorization processes. Júlio is also passionate about science education and equity, diversity, and inclusion in STEM. In addition to chemistry, Júlio is also passionate about Brazilian music and loves playing the guitar and singing; he also enjoys dancing, exercising, hiking, traveling, and learning new languages.

Kevin Jablonka (Swiss Federal Institute of Technology)



Kevin Jablonka was born in 1996 in a small town in Germany and spent a lot of time in the lab while in high school. He scored successes in student science fairs such as “Jugend forscht”, which he is now supporting as juror, or the International Conference of young scientists.
After high school, he obtained his bachelor’s degree in chemistry at TU Munich and then joined EPFL for his master’s studies in chemistry and an extended study degree in applied machine

learning. While at EPFL, he also spent half a year in the Quantum Chemistry and Modeling department of BASF in Tarrytown, New York.

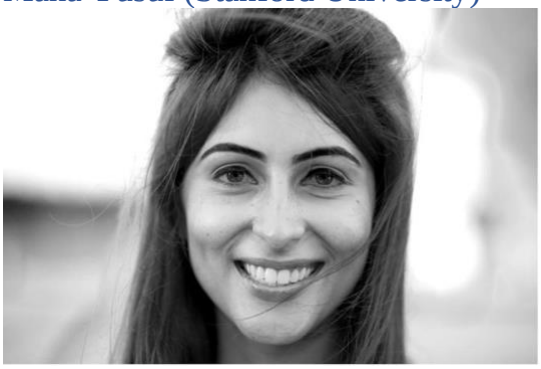
Kevin continued pursuing computational research during his Ph.D. in the group of Berend Smit, in which he focused on accelerating materials discovery across scales. got involved in the cheminfo ELN ecosystem. He also developed a toolbox for digital (reticular) chemistry and is co-leading the ChemNLP open science project, which aims at creating a foundational chemistry model. He now leads a research group at the Helmholtz Institute for Polymers in Energy Applications (HIPOLE).

Laura Bickerton (University of Edinburgh)



Laura Bickerton is currently undertaking a 3-year EPSRC funded postdoctoral position working within the Prof Paul Lusby group in collaboration with Prof Guy Lloyd-Jones (University of Edinburgh) and Prof Fernanda Duarte (University of Oxford) on metal-organic cages for catalytic applications. Laura completed her BSc (Hons) in Biochemistry and Biological Chemistry at the University of Nottingham in 2018, and subsequently completed her DPhil in Inorganic Chemistry at the University of Oxford under the supervision of Professor Matthew Langton and Professor Paul Beer. During her PhD research she designed, synthesised, and tested supramolecular anion receptors, capable of facilitating the transport of biologically relevant anions across cell membranes, including the incorporation of halogen bonding and stimuli-responsive anion transporters. Laura's PhD research resulted in 4 first author publications, 3 poster prizes at the RSC Macrocyclic and Supramolecular Chemistry Meeting (2019, 2021 and 2022), and subsequently being named as a CAS Future Leader in 2022. Additionally, Laura is a mentee within the WISC (Women in Supramolecular Chemistry) mentoring network and is a mentor in association with the Royal Society of Chemistry. Now at the University of Edinburgh, she is a member of the postdoctoral committee working to ensuring a supportive working environment for all postdocs.

Maha Yusuf (Stanford University)



Maha Yusuf is an incoming Presidential Postdoctoral Fellow at Princeton University. She is currently finishing up her PhD in chemical engineering at Stanford University. In her thesis research, Maha has used high-resolution neutron and X-ray imaging techniques to investigate Li plating characteristics on graphite electrodes during extreme fast charging of lithium-ion batteries. During her PhD, Maha has led extensive beamtime experiments at national and international neutron and X-ray facilities: National Institute of Standards and Technology Centre for Neutron Research (NIST), Oak Ridge National Laboratory (ORNL), and Paul Scherrer Institut (PSI).

For her work, Maha has won awards including the 2022 American Chemical Society CAS Future Leader; 2020 Stanford Distinguished Student Energy Lecturer; and 2020 and 2021 ECS and American Institute of Chemical Engineers Travel Grants. Maha holds an M.S. in chemical engineering from Stanford (2017) and a B.E. from the National University of Sciences and Technology, Pakistan (2013). Prior to Stanford, Maha worked on oil and gas rigs as a drilling engineer for Schlumberger in amazon rainforests in Colombia. To date, Maha has published six peer-reviewed journal articles and holds two patents.

Saigopalakrishna “Sai” Yerneni (Carnegie Mellon University)



Saigopalakrishna “Sai” Yerneni earned a Bachelors in Technology from Anna University (India) in 2012 followed by a Masters’ in Biomedical Engineering from Carnegie Mellon University (CMU), Pittsburgh in 2015. He then continued at CMU as a Dowd Fellow to pursue his PhD in Biomedical Engineering. His doctoral work focused on understanding extracellular vesicle

mediated cell-to-cell communication in health and disease. Specifically, he developed biofabrication technologies to study extracellular vesicle biology in regenerative medicine, cancer and infectious diseases. Currently he is a postdoctoral fellow at CMU and is working on advanced drug delivery technologies for nucleic acid drugs and gene therapies.

Silvina "Silvi" A. Di Pietro (Lawrence Livermore National Laboratory)



Silvina A. Di Pietro graduated from Florida International University (FIU) in the Spring of 2021 with a PhD in environmental chemistry. In June of 2021, she joined the Class of 2021 National Nuclear Security Administration (NNSA) Graduate Fellowship Program sponsored by Pacific Northwest National Laboratory (PNNL) as a post-doctoral scientist at Lawrence Livermore National Laboratory (LLNL). Upon successfully completing her 1-year post-doctoral appointment, she transitioned to a research position at LLNL. Currently, she is working under the supervision of Dr. Kiel Holliday within the Materials Science Division where she is working on developing additive manufacturing into a viable technology for nuclear applications.

During her graduate studies, she was a Department of Energy (DOE) Fellow at the Applied Research Center (ARC) and chemistry student within the Department of Chemistry and Biochemistry in FIU. At ARC, she assisted in development of ammonia gas injection for uranium remediation at the DOE Hanford Site in Washington State. Her dissertation project was titled *Uranium Fate and Mineral Transformations upon Remediation with Ammonia Gas*. The ammonia injection remediation technique has the potential to treat and sequester uranium, the major contaminant of concern released in the vadose zone at the Hanford Site. Silvina has been a recipient of the highly competitive 2022 CAS Future Leaders program sponsored by the American Chemical Society, the 2018 *Innovators in Nuclear Technology R&D* awarded by the DOE Office Nuclear Energy and 2016 *Roy G Post Scholarship* Foundation awarded by the Waste Management Symposia. Silvina's professional interests include water chemistry, geochemistry, climate change, and remediation of contaminated areas. Her hobbies are traveling, dancing, playing tennis, visiting art museums, and painting.

Stanna Dorn (California Institute of Technology)

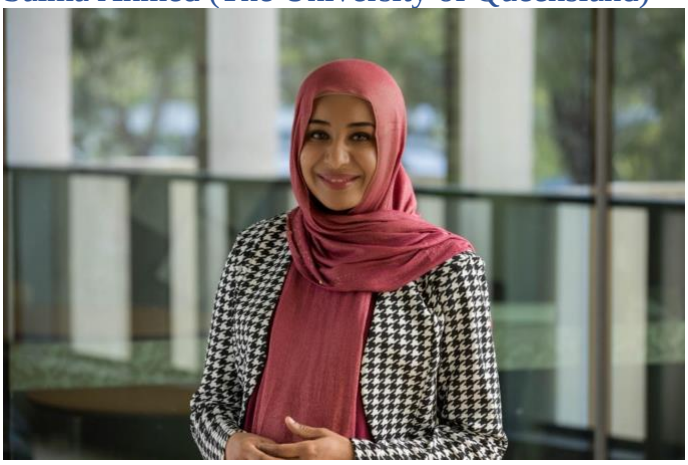


Stanna is an aspiring academic from Grayling, MI. She obtained a B.S. in chemistry and B.A. in music from Hope College, where she performed organometallic research with Prof. Jeffrey Johnson. She then received her PhD from Indiana University-Bloomington, where she worked with Prof. Kevin Brown on cooperative catalysis methods for alkene difunctionalization. Currently, Stanna is an NIH postdoctoral fellow in the Reisman lab at Caltech, working on total synthesis. Outside the lab, she enjoys reading, playing the flute, and proving that getting C's in organic as an undergraduate does not define one's future success in the field.

Olivia Harper Wilkins (NASA Goddard Space Flight Center)

Dr. Olivia Harper Wilkins, PhD, is a NASA Postdoctoral Program (NPP) Fellow in the Astrochemistry Laboratory at NASA Goddard Space Flight Center. In her current research, she primarily conducts cosmic ice experiments to understand how UV radiation affects interstellar, cometary, and planetary ices using the Sublimation of Laboratory Ices Millimeter/submillimeter Experiment (SubLIME). She also uses radio telescope observations to study the chemistry in high-mass star-forming regions. Dr. Wilkins earned her PhD in Chemistry at the California Institute of Technology (Caltech), where she conducted observational astrochemistry research as a National Science Foundation (NSF) Graduate Research Fellow. Before that, she earned a B.S. in Chemistry and Mathematics from Dickinson College and was a Fulbright Research Fellow in Germany. Besides research, Dr. Wilkins is involved in the NASA Goddard Association of Postdoctoral Scholars (NGAPS+) and is a coordinator for the NASA CONNECTORS (CONNECting high school students TO ResearcherS) program. She is also a Member-at-Large for the Maryland Section of the American Chemical Society (ACS) and an Affiliate of the National ACS Younger Chemists Committee (YCC). Dr. Wilkins was named a 2022 CAS (a division of the American Chemical Society) Future Leader and an AGU Science Activation (SciAct) Affiliate. She is passionate about science communication and using art to share science, something that culminated in her writing and illustrating *Astrochemistry* (2021) for the ACS's In Focus series. Dr. Wilkins enjoys traveling with her husband and son and spending time at coffee shops.

Salma Ahmed (The University of Queensland)



Salma Elgaili Ahmed is a final-stage PhD candidate of bioengineering under the supervision of Prof. Kristofer Thurecht at The University of Queensland, Australia. Her chemical engineering background combined with her MSc and PhD training inspired her to work in the interface of engineering and science to develop next generation tools for drug delivery. Salma has worked within highly interdisciplinary teams and developed myriads of nanocarriers for applications including precision cancer nanomedicine, diabetes treatment and miRNA delivery.

Salma is highly dedicated to her field. She has earned several accolades, including a CAS Future Leaders fellowship from the American Chemical Society, fully-funded MSc and PhD scholarships, and multiple other sci-com, travel and publication awards.

She is also a skilled communicator, where she has produced several high-impact publications and has presented her research at prestigious conferences and platforms. Salma is an active member of the wider community. She is an associate member of the Royal Society of Chemistry (AMRSC) and is dynamically involved in teaching and mentoring students in STEM fields. She was also a TEDxOmdurman Speaker, a TEDxUQ Curator and Mentor, and an accredited MHFAider.

As she approaches the end of her PhD, Salma is eager to embark on the next phase of her career. She is excited to apply her skills and knowledge to solve real-world problems in both academia and industry.